

Judicious Agent Theory

Andrew Pinchin

BA(Hons) LLB



Submitted for the degree of Doctor of Philosophy
ARU0360 Research in Philosophy

Monash University
Melbourne
Australia

September 26, 2012

Under the Copyright Act 1968, this thesis must be used only under the normal conditions of scholarly fair dealing. In particular no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

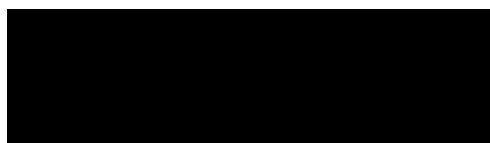
I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Abstract

This thesis is about moral facts and their relation to moral agents. It denies that moral facts are 'independent' of moral agents or explained by 'ideal' versions of those agents because the implications of these views are: (1) justifiably counter-intuitive; and (2) sufficiently implausible for independent reasons. Instead it is maintained that moral facts are explained by facts about what the particular moral agent in question could actually recognize and be motivated to do when reasoning as well as they could in the circumstances. Moreover, it is contended that this fact/ agent relation does not merely arise relative to specific moral agents in the sense in which relativity matters for moral facts in metaethics. Moral facts apply neutrally among moral agents to which they should apply.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and to the best of my knowledge it contains no material previously published or written by another person, except where due reference is made in the text.



Andrew Pinchin

September 26, 2012

Preface

This thesis is about the source of moral values and moral requirements. In particular, it is about whether this source could consist in facts in the world about what we should do and whether these facts can, in the end, be identified by a version of the scientific method. In light of two propositions that are somewhat familiar in metaethics, but not so often defended, this thesis contends that while moral facts exist in a certain sense, their existence is less robust than is sometimes supposed and unidentifiable by the scientific method.

An important part of this argument lies in taking the limitations of human agents seriously, at least where the extent of these limitations is stipulated to. In light of this emphasis, this thesis contends that moral values and requirements consist in facts about what it is actually possible for moral agents to decide and be motivated to do.

Upon starting out I did not anticipate defending such a conservative view, and specifically, defending a theory maintaining that there is a necessary relation between reasons for people to act and their being motivated to do so. As is sometimes the case, I came to do so in virtue of believing that far more could be said in favour of this view than is often granted. In therefore defending such a view, and especially the rather conservative line that I present, hopefully I have at least provided reason to judge it with greater earnest, even if writers disagree with it.

One point on style requires mentioning at the outset. Where I want to emphasise a word or expression already italicized as part of a quote or as a symbol or abbreviation at times, for ease of understanding, I distinguish my own emphasis with underlining rather than in footnotes.

Acknowledgements

I am very grateful for the help of my supervisor, Karen Green. Her timely comments always helped me to identify and ameliorate weaknesses in my work which would not be where it is if it were not for her time and effort. I also thank Toby Handfield for his comments on part of my work and participants at talks given on the material for their time and contributions.

Table of Contents

Abstract.....	iv
Declaration	iv
Preface.....	v
Acknowledgements	vi
1 REALISM & REASONS	1
I Moral Realism	1
A. Moral Impartiality: Independence	9
B. Moral Impartiality: Agent-Neutrality	14
C. Ideal Observer Theory	26
D. Natural Facts.....	27
II Justifying Reasons.....	30
A. Justifying Reasons	31
B. Explanatory Reasons	36
III Outlook.....	37
2 OPEN QUESTIONS & CONTEMPORARY ANSWERS.....	40
I The Open-Question Argument	43
II The Knowledge Argument.....	49
III Objection 1: Vicious Circularity	52
A. OQA	52
B. The Knowledge Argument	58
IV Objection 2: Invalid Inference	60
A. OQA	60
B. The Knowledge Argument	63
V Objection 3: Property Identification.....	66
A. OQA	66
B. The Knowledge Argument	73
VI Objection 4: Invalidity	74
VII Conclusion.....	75
3 MOTIVATION IN INTENTIONAL CONDUCT	77
I The Humean Theory of Motivation.....	80
II Relative & Contingently Neutral Explanations	88

III	The Last Word	98
IV	Morality and R&CN.....	109
VI	Conclusion & Outlook	123
4	JUSTIFYING REASONS & THE POSSIBILITY OF MOTIVATION	124
I	Williams' Internalism.....	127
	A. Kinds of Reasoning.....	130
	B. Cognitivism about Reasons	132
	C. Possibility	133
	D. Weak, Ignorant, Akritic & Strong Internalism.....	135
II	Externalism.....	139
III	The Conditional Fallacy.....	143
	A. Railton & Smith.....	145
	B. Johnson's Objection	146
	C. Answering Johnson's Trilemma	148
5	MODEST INTERNALISM I.....	155
I	External Reason Statements.....	158
	A. Consideration 1: Implausibility of Content.....	165
	B. Consideration 2: Rarity of Knowledge	169
	C. Consideration 3: Other Expressions	171
II	Intuitive Evidence	172
	A. Limitations on Reason Ascriptions.....	174
	B. Non-Rational Intervention.....	177
	C. Triviality.....	185
III	Williams' Argument from Parsimony	186
	A. Reply: Evaluations Entail Reasons	188
	B. Evaluative Judgments & Reasons.....	190
	C. Gibbard's Objection.....	192
6	MODEST INTERNALISM II	197
I	Justifying Reasons & Truth.....	200
II	Counterexamples	202
	A. Type 1: Williams, Shafer-Landau, Kahane & Markovits.....	203
	B. Type 2: Millgram & Schroeder.....	208
	C. Type 3: Hampton & Lillehammer.....	213
III	Holding A Responsible	214
	A. Holding A Responsible for not Developing her S	216

B. Holding <i>A</i> Responsible: Specific Deterrence	220
C. Holding <i>A</i> Responsible: General Deterrence	221
IV Theoretical Reasons	222
V Review & Outlook	224
 7 MORALITY/ REASON INTERNALISM.....	226
I Analysis of a Justifying Reason.....	227
II Morality/ Reason Internalism Disambiguated	229
III MF—R Internalism: Justification.....	234
IV Moral Facts & Recognition	242
 8 MORAL VALUES & AGENTS	245
I Response-Dependent Theories: Extension-Reflecting & Extension-Determining Accounts	246
A. Sensibility Theory.....	250
B. Indications of Substantive Moral Properties from Dimensions of Objectivity.....	254
C. Rigidity: Colours and Gustatory Properties	259
D. Rigidity: Moral Properties	267
E. Sensibility Theory Qualified	271
II Judicious Agent Theory	273
A. Objections.....	274
 Bibliography.....	284
Index	302

I

Realism & Reasons

Our enquiry begins by outlining a theory affirming the existence of moral facts, a *realist* theory of morality. Part of this process consists in identifying whether moral facts—whatever they turn out to be—are facts recognizable by science in some way, or, whether they are *natural*. Both of the terms ‘realism’ and ‘natural’ require explanation and I define each in Section I below. As the core argument against realism that I discuss in this thesis relies on different senses of one’s ‘reason for acting’, Section II in this chapter turns to distinguishing these senses. Accordingly, the aim of this chapter is merely to present the foundations for the rest of the thesis which is crucial given the present state of the art. With this in place, Section III surveys the structure and content of Chapters 2-8.

I Moral Realism

There is disagreement as to whether a generic definition of realism can be applied *mutatis mutandis* to various philosophical domains, so I outline the theory of moral realism we are concerned with independently of these considerations.¹ While *moral realism* is a term of art, there is no consensus as

¹ Several writers define ‘realism’ in a way that they consider to be applicable or loosely applicable across philosophical domains: Brink, ‘Moral Realism and the

to its definition, largely due to disagreement over which of a range of potential elements are essential to it, if any; what those terms mean; and whether their implications are agreeable.² It is not one of the aims of this thesis to argue that a specific definition is correct, if such exists, however given this present lack of consensus I at least explain why I use the terms I do. In this way, as R. M. Hare explains, even where there is disagreement over my definition, at least my approach may be understood.

In sketching a theory of moral realism (and realism generally) one is faced with determining the significance of a number of potential realist characteristics, including, but not limited to: cognitivism, success theory, independence, non-minimalism, causal efficacy and explanatory power. In identifying which of these characteristics are essential to moral realism I let two ideas guide my considerations. Firstly, in using the term ‘moral realism’ I mean a view that is opposed to individual and cultural metaethical relativism, though not necessarily circumstantial or temporal relativism and

Sceptical Arguments from Disagreement and Queerness’, 111; Brink, *Moral Realism and the Foundations of Ethics*, 15; Sayre-McCord, ‘The Many Moral Realisms’, 2; Haldane and Wright, ‘Introduction’, 3–4; Pettit, ‘Realism and Response-Dependence’, 588–9; Bloomfield, *Moral Reality*, ix–x; Oddie, *Value, Reality, and Desire*, 2–3; Miller, ‘Realism’, sec. Introductory matter; Brock and Mares, *Realism and Anti-Realism*, 2; Skorupski, *The Domain of Reasons*, 7. In contrast, Mark Eli Kalderon, Richard Joyce and Geoffrey Sayre-McCord criticize the need for a schematic definition. For example, Kalderon argues that a general definition of realism would have to abstract away from the different terms each domain employs and the worry then becomes that a general definition would be too general to suit our purposes. Joyce argues that it seems improbable that all outlooks called ‘realism’ have a shared core of commitments, such as, for example, the French realist art movement. He acknowledges that one may confine applicable domains to philosophically relevant subjects, but he argues that the point still remains. Kalderon, *Moral Fictionalism*, 100; Joyce, ‘Moral Anti-Realism’, sec. 1; Sayre-McCord, ‘Being a Realist About Relativism (In Ethics)’, 159.

² Peter Railton defines moral realism in virtue of 13 dimensions, noting that other proponents of moral realism may agree to more or less of the dimensions he signs onto and perhaps still be ‘realists’: Railton, ‘Moral Realism’, 164–5. Indeed, Hilary Putnam’s Internal Realism may be true at the level of metaethics: there are as many ways of describing realism available as there are conceptual schemes available to view the metaethical landscape: Putnam, *The Many Faces of Realism: The Paul Carus Lectures*, 17.

not necessarily pluralism. While it may be that, in defining moral realism, some academics are more concerned with criteria requiring robust metaphysics or knowledge of true moral judgments, I consider robust metaphysics and knowledge of true moral judgments important insofar as they provide grounds for opposing certain kinds of relativism.³ I am less concerned with circumstantial and temporal relativism only because differences in moral evaluations due to differences in circumstances and times are more intuitively acceptable for moral realism and a test of intuitiveness suffices for the purpose of merely defining realism.⁴

The second guiding idea is that moral realism asserts that certain morally relevant facts or properties exist in something more than a merely nominal or minimalist sense. By a 'nominal' or 'minimalist' sense I mean the sense in which the fact or property is reducible to a proposition and hence, a sense lacking metaphysical commitment. Commitment to moral facts or properties in a more substantive sense than this is, as I understand the name, what makes the theory *realist*, just as realism about logs and kangaroos is committed to there being logs and kangaroos and not merely true propositions about them. While I take these two guiding ideas to be important, I am more relaxed about the theory's adherence to other criteria sometimes introduced. For this reason, as long as the theory is consistent

³ I am not alone in this regard, C. S. Jenkins commenting, for example, that it is common to define realism as opposition to some kind of relativism, although writers also often disagree: Jenkins, 'What Is Ontological Realism?', 883. For example, Mark Eli Kalderon considers that moral realism is compatible with relativism because moral facts can be relations. In this regard, Kalderon's point is analogous to the point that the tenseless theory of time maintains that relations can explain temporal difference. Kalderon, *Moral Fictionalism*, 114; Moore, 'Time and Well-Being', 86. Geoffrey Sayre-McCord also defends the view that realism is compatible with relativism and Bruce W. Brower agrees with this: Sayre-McCord, 'Being a Realist About Relativism (In Ethics)'; Brower, 'Dispositional Ethical Realism', 225 n5.

⁴ Jonathan Dancy defends a theory that he describes as moral realism and which nevertheless allows for rather extensive circumstantial relativism. Dancy, *Ethics Without Principles*. Peter Caws argues that moral realism is compatible with a form of temporal relativism: Caws, 'Ethics and Temporality: *When Are Moral Propositions True?'

with these two guiding ideas, I interpret moral realism liberally so as to take into account as wide a range of metaethical theories as possible.

Accordingly, for a metaethical theory to be a version of *Moral Realism*, it is a necessary and sufficient condition that the theory maintains that the following five defining propositions are true:

1. **Truth-Apt Propositions:** The contents of actual and possible moral evaluations bear a semantic relation to propositions ('moral propositions') which are truth-apt.
2. **Representationalism:** At least in some cases where Truth-Apt Propositions is true, moral propositions are true only if they correctly represent certain facts and they are false only if they do not correctly represent those facts. (*Denial of 1 and/ or 2 entails Non-Factualism*)
3. **Propositional Truth:** There is at least one moral proposition that is true in virtue of representing a certain fact(s). (*Denial of 3 entails Extreme Moral Skepticism*)⁵
4. **Non-Minimalism:** In propositions 1-3, moral propositions, truth values, references and facts are not understood according to minimalist theories. (*Acceptance of 1-3 and denial of 4 permits: Idealism, Cognitive Irrealism, and Quasi-Realism*)
5. **Moral Impartiality:** In propositions 1-3, either:
 - a) true moral propositions (and their contents), truth values, references and facts (and their contents) exist independently of moral agents; or
 - b) where a) does not apply to a given entity, any relation between the entity and moral agents is agent-neutral. (*Denial of 5 permits Metaethical Individual Relativism and Metaethical Cultural Relativism*)

⁵ I am following Aaron Zimmerman in regarding Extreme Moral Skepticism as the view that we cannot have *justified moral beliefs*, compared to Moderate Moral Skepticism, the view that we cannot have moral *knowledge*: Zimmerman, *Moral Epistemology*, 42.

Natural Moral Realism is defined by an additional, sixth proposition:

- 6. Natural Facts:** In Truth-Apt Propositions and Representationalism, the facts in question are natural facts. (*Acceptance of 1-5 and denial of 6 entails either Non-Natural Moral Realism or Supernatural Moral Realism*)

Some clarifications are required. Firstly, Truth-Apt Propositions and Representationalism do not provide that the facts in question must be *moral* facts. I allow this possibility because I need not consider the issue and at least one writer argues to the contrary.⁶ Truth-Apt Propositions and Representationalism are therefore intentionally silent on the matter and, hereafter, my use of the term ‘moral fact’ applies to facts determining the truth or falsity of moral propositions—whether or not such facts are otherwise distinctively ‘moral’ facts.

Secondly, Propositional Truth raises one of two key issues dividing writers defining moral realism: one epistemic, one metaphysical.⁷ The key epistemic issue is whether an epistemic requirement also arises, such as a moral agent being capable of recognizing the moral fact. Specifically, some writers consider that a version of Cognitivism and Success Theory below should replace proposition 3, Propositional Truth.

- 3a*. Cognitivism:** At least in some cases where Representationalism is true, moral evaluators believe that the content of their moral evaluation is true in light of believing that a fact(s) corresponding to the evaluation obtains. (*Denial of 3a* entails Non-Cognitivism. Acceptance of 1 & 2 and denial of 3a* entails Hermeneutic Fictionalism*)

⁶ Dyke, ‘What Moral Realism can Learn from the Philosophy of Time’.

⁷ Indeed, both of these issues pervade definitions of realism generally. For example, Crispin Wright discusses Michael Dummett’s recognition-transcendent truths and states of affairs whose obtaining or not is beyond our detection. Wright, *Truth and Objectivity*, 4–5.

3b*. Success Theory: At least in some cases where Representationalism and Cognitivism is true, at least one moral proposition is true. (*Acceptance of 1, 2 and 3a* and denial of 3b* entails Error Theory or Moderate, if not Extreme, Skepticism*)

To this end, Stuart Brock and Edwin Mares argue, firstly, that an object that exists outside of our light cone is no less real for being cognitively inaccessible.⁸ Secondly, they maintain that an epistemic requirement does not fit well with a requirement of independence: given that our minds could grow stronger or weaker, moral truths should not be said to vary in correspondence with changes of this nature.⁹

Contrary to this, Crispin Wright writes:

There are, no doubt, kinds of moral realism which do have the consequence that moral reality *may* transcend all possibility of detection. But it is surely not essential to any view worth regarding as realist about morals that it incorporate a commitment to that idea.¹⁰

It is also clear that some writers include an adherence to some version of moral cognitivism in their definition of moral realism.¹¹ In addition, it might

⁸ Brock and Mares, *Realism and Anti-Realism*, 6–7.

⁹ Ibid., 7. Terence Cuneo and Guy Kahane also argue, respectively, that practical truths and moral facts may not necessarily be known: Cuneo, 'Are Moral Qualities Response-Dependent?', 583; Kahane, 'Must Metaethical Realism Make a Semantic Claim?'. François Schroeter defines moral realism as allowing that moral facts may outstrip our knowledge of them without arguing for this claim: Schroeter, 'Reflective Equilibrium and Antitheory', 128.

¹⁰ Wright, *Truth and Objectivity*, 9 (emphasis added).

¹¹ William J. Fitzpatrick states that Geoffrey Sayre-McCord's slightly revised criteria comprise two claims that nearly all moral realists accept, at least as necessary conditions. FitzPatrick, 'Recent Work on Ethical Realism', 746; FitzPatrick, 'Robust Ethical Realism, Non-naturalism, and Normativity', 161. Sayre-McCord's initial two criteria arise in: Sayre-McCord, 'The Many Moral Realisms', 3. His slightly revised version appears in: Sayre-McCord, 'Moral Realism [Oxford Handbook]', secs. 1–2; Sayre-McCord, 'Moral Realism [SEP]', sec. Introductory matter. The following writers, for example, regard moral realism as including a cognitivist claim of some kind: McNaughton, *Moral Vision*, 39; Haldane and Wright, 'Introduction', 3–4; Skorupski, *The Domain of Reasons*, 428. Yet I doubt that nearly all writers regard moral realism as involving an epistemic requirement: as noted above, several writers

be argued that, if we could not cognitively access facts that render our moral evaluations true, and if we could not believe that our moral evaluations are true in virtue of those facts then, there is no reason to posit them at all.

In light of this last concern I hasten to mention that the versions of Cognitivism and Success Theory in 3a* and 3b* are more modest than some versions. Cognitivism is a little demanding in light of requiring that at least in some cases moral agents believe that the content of their moral evaluation is true in light of believing certain facts. It is somewhat demanding because if moral evaluations consist in beliefs in some cases, it is a short step to claiming that this is always the mental state of moral agents when they make moral evaluations. Arguably, however, in the context of defining moral realism it is still a relatively modest claim given that, so far as I know, no moral realist denies that moral agents sometimes believe that their moral evaluation is true in light of a certain fact(s). This is the force of Wright's point.

Success Theory in 3b* is even more modest given that it only requires that one such moral evaluation is true. It therefore does not require that moral agents can *always* cognitively access the fact(s) that would allow them to make the right moral evaluation. Significantly, this means that it does allow for the existence of recognition-transcendent moral facts. It only requires that at least one such fact(s) is not recognition-transcendent and is accessed by a moral agent at one point in the history and future of humanity. I also know of no moral realist who denies this.

In light of these considerations, a definition of moral realism given by substituting 3a* and 3b* for proposition 3 (Propositional Truth) would adequately represent writers' present definitions of moral realism, perhaps if

define moral realism independently of the cognitive accessibility of moral facts or true moral propositions.

this is taken as including commitments that they implicitly take to follow from it. A definition in these terms also, conveniently and importantly, identifies the relationship of moral realism to various other irrealist theories: moral noncognitivism, hermeneutic moral fictionalism, moral error theory and moderate moral skepticism. Yet, for all this, Brock and Mares are right; therefore, a preferable and more cautious statement of moral realism will exclude any such epistemic requirement. Moreover, definitions of theories in philosophy generally are better off for not unnecessarily closing off conceptual space and, by defining moral realism so that it must include moral cognitivism, we would be unnecessarily closing off the conceptual possibility of some breed of noncognitive realism. On its face, such a theory faces certain difficulties. Yet this alone does not justify defining away such a position, especially when nothing is lost in defining realism more modestly given that it can always be supplemented by Cognitivism and Success Theory, or versions thereof, anyway. I therefore define moral realism in virtue of Propositional Truth and not 3a* and 3b*. So understood, Moral Realism remains neutral on the question of whether any true moral proposition and the facts that make these propositions true are cognitively accessible. They may or may not be, depending on whether one also accepts a version of Cognitivism and Success Theory.

Only a brief comment on proposition 4 (Non-Minimalism) is required, yet the terms of proposition 5 (Moral Impartiality) require clarification given that the primary focus of this thesis lies in an argument disputing the truth of it. Non-Minimalism is enshrined as my second guiding idea and it is not particularly, if at all, controversial.¹² I therefore turn to considering the terms

¹² For example: Skorupski, 'Irrealist Cognitivism', 456; Skorupski, 'The Ontology of Reasons', 118–9; Skorupski, *The Domain of Reasons*, 7, 404–5; Cuneo, *The Normative Web*, 27; Robertson, 'Introduction: Normativity, Reasons, Rationality', 13.

‘independently’ and ‘agent-neutral relation of dependence’ in Moral Impartiality.¹³

A. Moral Impartiality: Independence

In addition to propositions 1-4, it is common for writers to require that moral propositions, truth values, references or facts, (hereafter for convenience I just mention truth values), are objective or independent in some sense.¹⁴ However, it is controversial whether the meaning of the terms ‘objective’ and ‘independent’ can be stated both precisely and without counter-intuitive implications and this concern therefore comprises the key metaphysical issue dividing writers’ definitions of moral realism.¹⁵ Sometimes this criterion is

¹³ I use the term ‘Moral Impartiality’ because ‘moral objectivity’ and ‘moral absolutism’ are already used in several different ways, ‘moral egalitarianism’ and ‘moral fairness’ each have unwanted connotations, and ‘moral neutrality’ too closely resembles ‘agent-neutrality’.

¹⁴ Sturgeon, ‘What Difference Does It Make Whether Moral Realism Is True?’, 117; Dancy, ‘Two Conceptions of Moral Realism I’, 167–8; Boyd, ‘How to Be a Moral Realist’, 182; McNaughton, *Moral Vision*, 44–5; Tännsjö, *Moral Realism*, 111; Pettit, ‘Realism and Response-Dependence’, 590; Bloomfield, *Moral Reality*, ix; Haldane and Wright, ‘Introduction’, 3; Brower, ‘Dispositional Ethical Realism’, 238; Shafer-Landau, *Moral Realism*, 2, 15, 170 n7; Korsgaard, ‘Realism and Constructivism in Twentieth-Century Moral Philosophy’, 303; Huemer, *Ethical Intuitionism*, 4–6, 10; Oddie, *Value, Reality, and Desire*, 2–3; Loux, *Metaphysics*, 263; Street, ‘A Darwinian Dilemma for Realist Theories of Value’, 111–2; Brock and Mares, *Realism and Anti-Realism*, 113; FitzPatrick, ‘Robust Ethical Realism, Non-naturalism, and Normativity’, 166; Miller, ‘Realism’, secs. Introductory matter, 6, 7; Jenkins, ‘What Is Ontological Realism?’, 881, 883; And, in regard to normativity in general (at least): Frankfurt, *Taking Ourselves Seriously and Getting It Right*, 33; Wedgwood, *The Nature of Normativity*, 1–2; Skorupski, *The Domain of Reasons*, 7, 8, 404–5. David O. Brink and Stephen Finlay also make use of a criterion of independence, albeit with reservations. For example, Finlay distinguishes different faces of realism, at times on the basis of this criterion: Brink, *Moral Realism and the Foundations of Ethics*, 15; Finlay, ‘Four Faces of Moral Realism’, 822, 829.

¹⁵ Criticism of the ‘independence’ criterion comes from: Putnam, *The Many Faces of Realism: The Paul Carus Lectures*; Alston, ‘Introduction’, 1; Joyce, ‘Moral Anti-Realism’, sec. 5; And, to an extent: Sayre-McCord, ‘Moral Realism [SEP]’, sec. 5. Compare John Skorupski’s argument: ‘although the thesis that some facts are cognitive independent may not be clear,’ it should not be denied that there is any significant sense in which it is true. ‘That there are substantial facts, [e.g. that it is getting dark] and that they are in some significant sense independent of our

employed to stave off the threat of idealism (the view that all entities in a given discourse are mental entities);¹⁶ however, idealism is precluded by the requirement that morally relevant facts and properties exist in a substantive or metaphysically committed sense.

Independence is also sometimes used to ward off the threat of agent-relativism.¹⁷ If the truth observed is thought to exist independently of the agent in some sense, then it cannot be said that its existence and content varies among agents properly observing it. However, if agent-relativism is the motive for introducing a criterion of independence, it is more appropriately answered by asserting that either the truth in question exists independently of agents *or* that the dependence relation that obtains is agent-neutral (the second disjunct in Moral Impartiality). Thus, while I define 'independence' below, even if the truth in question is not independent of the agent (in the sense to be explained) the theory may still be a variant of Moral Realism if the truth bears an agent-neutral relation to agents. I consider the terms 'independent' and 'agent-neutral' in turn.

There are two senses of independence in regard to the truth or falsity of a proposition that can be quickly set aside. Firstly, where an evaluation refers to an agent's mental state or her mind, then the truth or falsity of the proposition in question will clearly depend on that agent's mind. For this reason, it is sometimes objected that realists will not consider it desirable to claim that the agent's mental state or mind is not real, so the criterion of

knowledge of them, is too much a part of common sense to be baldly rejected': Skorupski, *The Domain of Reasons*, 8–9.

¹⁶ For example, Graham Oddie notes that traditionally, mind-dependence is the hallmark of idealism and Panayot Butchvarov implies that the requirement of independence arose to deal with idealism from Berkeley: Oddie, *Value, Reality, and Desire*, 14; Butchvarov, 'Saying and Showing the Good', 147.

¹⁷ Fitzpatrick makes this point using Bernard Williams' theory as an illustration of a theory that we would not want to identify as being realist. FitzPatrick, 'Robust Ethical Realism, Non-naturalism, and Normativity', 165–6.

independence is flawed.¹⁸ Yet, as several writers comment, trivial dependencies, such as this, are not what is at issue between realists and anti-realists.¹⁹

Secondly, realists are not concerned to assert that the truth values of propositions are *causally* independent of agents.²⁰ A finished painting only exists because of the touch of the artist's hand and similarly, the evilness of Wilson's dropping a match on the line of gasoline leading to the oil refinery is causally dependent on Wilson's deed. Realism can accept that the truth values of propositions are dependent on agents in this sense because the existence of the painting and the evilness of Wilson's act is not thereby 'dependent' on an agent in the sense realists are concerned to maintain and anti-realists deny.

Significantly more common is the claim that one thing, the blue colour of a blue and red golf ball, for example, is independent of another thing, the existence of human beings or creatures with perceptual capacities similar to our own, for example, if and only if there is a possible world in which the ball exists as blue and red but in which human agents do not. There are a number of criticisms of the use of such 'modal independence' to define realism. One problem is that the assertion that 'a given entity exists in a possible world in which humans or rational agents do not exist' fails to threaten anti-realism. Anti-realists can maintain that in asserting that the relevant entity exists in a possible world wherein rational agents are absent, one is only asserting that the entity exists because one *conceives of it* as

¹⁸ This is one of Brink's worries. Brink, *Moral Realism and the Foundations of Ethics*, 15.

¹⁹ Jenkins, 'Realism and Independence', 199 col. 2; Loux, *Metaphysics*, 263; Miller, 'Realism', sec. 1; Cuneo, *The Normative Web*, 46–7.

²⁰ Brink, *Moral Realism and the Foundations of Ethics*; Jenkins, 'Realism and Independence', 199 col. 1–2; Joyce, 'Moral Anti-Realism', sec. 5.

existing in another possible world absent rational agents.²¹ We revisit this concern in Chapter 8.

More problematic, however, is an objection that can be abstracted from the literature on the 'essence' of a thing. Often definitions of the essence of a thing rely on the modal claim that: the thing's essence consists in those properties that it is impossible for the object to exist without.²² However, it has been observed that this claim entails the result that, for example, in all possible worlds in which Humphrey Bogart exists, it is true that:

Humphrey Bogart is a member of the set of 'Humphrey Bogart and all the irrational numbers'.

It is impossible for Humphrey Bogart to exist without this proposition being true. Yet, the truth of this proposition is not part of the essence of Humphrey Bogart. The same point applies to definitions of 'independence' in modal terms and, accordingly, this sense of independence is sufficient but not necessary: there are cases where we are willing to classify one thing, A, as being independent of another thing, B, which the requirement of modal independence does not allow for.

Rather, in claiming that the truth values of propositions are independent of moral agents, C. S. Jenkins has the right of the matter by regarding truth values to be independent of rational agents in the sense that they are 'essentially independent' of those agents. By essential independence, Jenkins means that it is not a condition of the existence of truth values that human

²¹ For further arguments against modal independence see: Dancy, 'Two Conceptions of Moral Realism I', 167–8; Fine, 'Essence and Modality: The Second Philosophical Perspectives Lecture'; Jenkins, 'Realism and Independence'.

²² For example: Wedgwood, 'The Essence of Response-Dependence', 46.

agent(s) exist.²³ Thus, for example, a realist about mathematical truths and mathematical objects may claim that the existence of mathematical truths is independent of human beings in the sense that the existence of such truths is not conditional on the existence of human beings.

Essential independence is a necessary condition of modal independence but the reverse is not true. It may be that some entity, ϑ , exists that is *essentially* independent of another entity, τ , in possible world w even though there is no possible world in which ϑ exists and τ does not. We saw this in regard to Humphrey Bogart. Similarly, suppose, following Jenkins, that in every possible world in which I exist, it is true that $3+2=5$. Yet, it is not a necessary condition of *my* existing in any possible world that it is true that $3+2=5$. Rather, we think that $3+2=5$ is true in those worlds in which I exist (and in other possible worlds) for *other* reasons.²⁴ Thus, even though there is no possible world in which I exist and in which it is false that $3+2=5$, I still enjoy an existence that is essentially (but not modally) independent of the mathematical truth of $3+2=5$ since this truth's obtaining is not a condition of my existence.

Similarly, we can say that the moral evaluation 'cruelty to animals is morally evil' has a semantic relation to a proposition, p , which has a truth value (say, true). The truth of p will be essentially independent from human agents existing if and only if it is not a condition of the truth of p that human agents exist. It may be that p is true in a possible world in which humans do not exist but in which other rational agents exist. If so, then the truth of p is also modally independent of the existence of human agents. But modal independence is not required for the truth value of p to be essentially

²³ Jenkins, 'Realism and Independence', 200 col. 1 – 202 col. 2. Jonathan Dancy endorses a similar definition: 'real properties are those which are not constituted by the availability or possibility of a characteristic human response.' Dancy, 'Two Conceptions of Moral Realism I', 168.

²⁴ Jenkins, 'Realism and Independence', 201 col. 1.

independent of the existence of human agents. Thus, even if in every possible world where p is true, human agents also exist, it may still be that p 's being true is essentially independent from human agents existing because it may be that it is no part of the conditions for the existence of the truth of p that human agents exist.²⁵

The explanation just given in terms of essential independence from human agents' existence clearly maintains the analogy to the case of mathematical truths, but for a definition of *moral* realism we need not suppose that the independence relates to *the existence* of human or moral agents. We may instead make the weaker claim that it is no part of the conditions for the existence of the truth of moral proposition p that: **a)** actual moral agents genuinely believe p ; **b)** actual moral agents endorse standards that determine that p is true; or **c)** that actual or hypothetical moral agents would believe p or endorse standards that determine that p is true.

In light of this, hereafter by 'independence' and 'dependence', I mean, respectively, essential independence and essential dependence. With this understood, I consider the modifier 'agent-neutral'.

B. Moral Impartiality: Agent-Neutrality

Since inception, the terms agent-neutral and agent-relative have been employed with various meanings and in regard to an array of objects.^{26 / 27} In

²⁵ This point about the problem of modal independence generalizes to all cases of necessitation and so we shall be revisiting it.

²⁶ Different definitions of the terms are given by: Parfit, *Reasons and Persons*, 143; Nagel, *The View From Nowhere*, 152–3; Pettit, 'Universalizability without Utilitarianism', 75; Dancy, 'Agent-Relativity -- The Very Idea', 234; Hurka, 'Moore in the Middle', 611. Although, notably Michael Ridge argues that Parfit and Nagel's definitions come to the same test. Ridge, 'Reasons for Action: Agent-Neutral vs. Agent-Relative', sec. 1.

introducing my account of the distinction it will be convenient to discuss each term in regard to justifying reasons to act because some of the more important work on the distinction arises in this context and because we shall nevertheless be concerned with agent-relative and agent-neutral reasons in Chapters 3-7. The same distinction can then be applied to truth values. I explain what I mean by a justifying reason in more detail below. Suffice it to say for now that a justifying reason for an agent, *A*, to perform some act, Φ , is a reason that *A* is to some extent justified (or thinks she is justified) in Φ -ing for. The presence of this reason does not mean that *A* ultimately *does* Φ , only that she is at least to some extent justified in doing so (or thinks she is to some extent justified in doing so).

For the purpose of defining agent-relativity, consider relativity in general. To say that something (*X*) is true relative to something else (*Y*) is to say that given *Y*, then *X* is true. Thus, to say ‘the abandoned church is more than 20 kilometres away’ (*X*) is to say ‘the abandoned church is more than 20 kilometres away relative to a certain standard’, say ‘the standard: where Valerie Frost is at time *t*’ (*Y*). Having said this, it may or may not be that *X* is true relative to a different standard. Perhaps, for example, Jaegar Loewe is standing in front of the abandoned church at *t*, in which case, relative to the standard ‘where Jaegar Loewe is at time *t*’ (*Z*), *X* is false. Thus, in general, for propositions that are true or false relative to something else, the proposition is neither true nor false *simpliciter*. The proposition is only true or false relative to some specified standard.

In one sense of ‘agent-relativity’, agent-relative reasons also operate in this way. If we hold fixed a certain standard (in this case, a rational agent, *A*),

²⁷ The literature discusses agent-relativity in regard to, for example: reasons, values, theories, aims, rules, principles, good (as a value), obligations, rights and duties, norms and questions, intuitions. But it may be that the term cannot take just any object: Schroeder, ‘Teleology, Agent-Relative Value, and “Good”’.

reasons r & s may exist for A . However, in regard to a different rational agent, B , it may be that reasons r & s do not exist and if this is so, it is not true that r & s exist *simpliciter*. Rather, r & s exist as reasons relative to a certain rational agent A (or in some cases, a proper subset of rational agents that includes A). Furthermore, even if reasons r & s do exist for A and all other rational agents who presently exist, it may still be true that r & s only exist *relative to all rational agents presently existing*. Usually, however, we are concerned with all possible rational agents because, other things being equal, limiting our focus to presently existing rational agents is arbitrary. Thus, in such circumstances it is only if reasons r & s exist for all actual and possible rational agents that it is true that r & s are non-relative or *agent-neutral* in regard to rational agents.

However, where all that is said is that ' r & s are reasons for A ', then it may or may not be that r & s also exist for all other rational agents and, hence, it may or may not be that r & s exist relative to A . I distinguish such reasons as what Mark Schroeder calls '*agent-relational* reasons'.²⁸ An agent-relational reason is a reason for a rational agent or a proper sub-group of rational agents. Thus, the reason corresponding to each of the first two reason statements below are agent-relational, while the reason corresponding to the third is not:

1. That Norman enjoys driving fast cars is a reason for Norman to drive a fast car on Saturday.
2. That Lera has an interest in living well is a reason for her to avoid excessive exposure to (ionizing) radiation.
3. There is a reason to support the Children's Medical Research Institute if one can.

²⁸ Schroeder, *Slaves of the Passions*, 16–7 (emphasis added). John Skorupski also makes this point, observing that an 'agent-relative' reason-predicate may express an agent-neutral reason if the reason could also be expressed by an agent-neutral reason-predicate. Skorupski, 'Agent-Neutrality, Consequentialism, Utilitarianism... A Terminological Note', 51; Skorupski, *The Domain of Reasons*, 60–1.

The reasons referred to in the first two reason statements are reasons for particular rational agents, Norman and Lera respectively, whereas the third is purportedly a reason for all rational agents (or all actual, rational human agents). Nevertheless, the reasons described by statements 1 and 2 will also differ if it turns out that all rational human agents have Lera's reason but not Norman's (or *vice versa*). If this is the case then Lera's reason will be agent-relational *and* a reason for all rational human agents, while Norman's reason will be agent-relational and agent-relative at *t* (for all rational human agents).

For *some* types of considerations (such as moral considerations), that give rise to both agent-relational reasons and agent-neutral reasons for members of a certain sub-group (such as rational human agents), there is disagreement about whether the agent-relational reason or the agent-neutral reason has explanatory priority over the other.²⁹ The issue is important. However, we shall see at the end of Chapter 3 that the argument of this thesis is actually consistent with either view.

Specifically then, I define agent-relational reasons as follows:

Agent-Relational Justifying Reason: A reason is *agent-relational* if and only if it contains an essential pronominal back-reference (or nominal reference) to the agent or a proper sub-group of all actual and possible rational agents when fully specified.

(The reason may also refer to the agent's circumstances and a certain time, but since relativity to circumstances and time is always present, I leave this aside).

²⁹ For example, John Skorupski defends the priority of agent-neutral reasons and, relatedly, David Alm argues that agent-neutral value has priority over relative value. Skorupski, *The Domain of Reasons*, sec. 13.1; Alm, 'An Argument for Agent-Neutral Value'. In contrast, Mark Schroeder and Toni Rønnow-Rasmussen maintain that agent-relational reasons have explanatory priority. Schroeder, 'Reasons and Agent-Neutrality'; Rønnow-Rasmussen, 'Normative Reasons and the Agent-Neutral/Relative Dichotomy', 231.

This definition adapts a version of Philip Pettit's definition of *agent-relative* reasons:

An agent-relative reason is one that cannot be fully specified without pronominal back-reference to the person for whom it is a reason. It is the sort of reason provided for an agent by the observation that *he* promised to perform the action in prospect, or that the action is in *his* interest, or that it is to the advantage of *his* children. In each case the motivating consideration involves essential reference to him or his.

An agent-neutral reason is one that can be fully specified without such an indexical device.³⁰

Pettit does not explain what 'fully specifying' a reason consists in aside from the implications we can draw from his illustrations, however he later comments that it consists in '*reason-supplying considerations*': 'considerations which must be taken to have borne rationally upon the formation of the agent's choice'.³¹ Pettit distinguishes these considerations from '*reason-supposed considerations*': considerations that must be supposed to exist in order for the reason statement to be the agent's own.³² He considers that only reason-supplying considerations form part of the agent's fully-specified reason. Thus, in Pettit's example, the reason-supplying consideration for an agent's helping an old lady may be that 'the old lady needs help'. In this case it is true that this consideration only influences the reason for which the agent acted insofar as it is true that the agent *believes* that the old lady needs help and the agent has the reason in question only insofar as he *wants* to help the old lady in such a case. But according to Pettit, the agent's belief 'that the old lady needs help' and her desire to help the old lady are reason-supposed considerations and are not part of the fully-specified reason.

My concern here, however, is for how precisely Pettit distinguishes between considerations that are reason-supplying and considerations that are reason-

³⁰ Pettit, 'Universalizability without Utilitarianism', 75 (emphasis present).

³¹ Ibid., 78.

³² Ibid., 79–80.

supposing. To this end, we must interpret the ‘considerations which must be taken to have borne rationally upon the formation of the agent's choice’ from the third-person perspective so that the agent need not have consciously taken these considerations into account. Without this interpretation the full specification of Norman’s reason to drive a fast car might be ‘to feel exhilarated’. In this case Norman’s *desire* to do drive the car does not form part of his conscious reason for doing it; rather it lies in the background.³³ But, according to Pettit’s test, Norman’s reason for driving a fast car must then be agent-neutral and this is implausible. The ‘considerations which must be taken to have borne rationally upon the formation of the agent's choice’ must therefore be interpreted more broadly than this. In what follows, I interpret it as including the considerations which we would cite from the third-person perspective, were we pressed for further justification for thinking that the reason in question exists for the agent. On this basis, Peter has a reason to help the old lady because she needs help and anyone who was at Peter’s location at the time would also have this reason. In contrast, Norman has a reason to drive the car because it is exhilarating, although anyone who was in Norman’s location at this time may not also have this reason.

Despite this, we cannot employ Pettit’s test as a test for agent-*relative* reasons because it fails to account for the possibility that what would be an ‘agent-relative reason’ on Pettit’s test is shared by everyone. Thus, for example, a full specification of Lera’s reason is: ‘I should avoid excessive exposure to radiation today because I have an interest in living well’. This reason includes a pronominal back-reference to Lera, and so, according to Pettit’s test, it is an agent-relative reason. Of course, however, we know from our discussion above that Lera’s reason may plausibly be considered to be a reason that everyone has, demonstrating that Pettit’s test cannot distinguish

³³ On this point see: Schroeder, *Slaves of the Passions*, chap. 2.

agent-relative reasons. It may be argued that the relevant pronominal reference in question is *inessential* to the statement of her reason because her reason can be fully-specified as: ‘everyone who has an interest in living has a reason to avoid excessive exposure to radiation’. Stated in this way, Lera’s reason would not involve a pronominal or nominal back-reference to Lera and it would therefore be an agent-neutral reason on Pettit’s test. But it is difficult to believe that this specification of Lera’s reason is what influences Lera’s decision to avoid excessive exposure to radiation. Rather, it is more plausible that Lera thinks that she should avoid excessive radiation because *she* has an interest in living well. Pettit’s test can therefore be applied to determine if a reason is agent-relational but not agent-relative and I use it for this purpose.

However, Pettit’s definition has been objected to on two grounds that would also affect my use of his formulation for agent-relational reasons if the objections were sustainable. Firstly, Michael Ridge argues that Pettit’s formulation commits one to a theory of the ontology of reasons that precludes reasons from being irreducible facts.³⁴ The problem is that if reasons are irreducible facts, then it appears that one must be committed to there being irreducibly *indexical* facts on Pettit’s formulation, yet this is controversial.

Ridge’s criticism would also apply to Pettit’s test as a test for agent-relational reasons. However, the above formulation avoids this concern because it allows that the reason relation that includes an agent as one of its relata refers back to the agent by her name rather than by a pronoun. In this way, the fact need not exemplify a property of identification and the formula does not preclude a particular ontological view of reasons.

³⁴ Ridge, ‘Reasons for Action: Agent-Neutral vs. Agent-Relative’, sec. 2.

Secondly, this definition of agent-relational reasons might be considered to be vulnerable to a criticism Jonathan Dancy puts to Thomas Nagel's definition of agent-relativity.³⁵ Dancy argues that an agent can have an agent-relative reason that is not indicated by a pronominal back-reference to the agent:

I have a reason to attend the conference rather than donating this money to famine relief because it will make it more probable that someone will find a cure for cancer.

We can see that this counterexample is potentially applicable to Pettit's formulation and also my definition of agent-relational reasons. Yet, the reason statement does include a pronominal back-reference to the agent: it is just that it appears as the first word in the reason statement. In this regard, Dancy's reason statement could also be rephrased so that the reference is more prominent.

However, this observation may not address Dancy's underlying concern because he also maintains that reasons that derive from an agent's personal projects are not agent-*relative* as this would introduce one thought too many into the agent's deliberations.³⁶ This is right, and it contrasts with the case of Lera above where we said that it is implausible to consider that her reason to avoid excessive exposure of radiation is 'that everyone has a reason to avoid doing so'. Conversely, it is sometimes implausible and undesirable for a rational agent to consider whether a purported reason for her to act in some way promotes one of *her* own personal projects before determining that the

³⁵ Dancy, 'Agent-Relativity -- The Very Idea', 236.

³⁶ It is curious that this point is potentially in tension with Dancy's own view of the distinction between agent-relative and agent-neutral reasons. Dancy argues that the distinction is to be understood intuitively by the thought that: I have a personal project that is very important to me but which I also know that in some sense it does not matter very much whether I succeed or fail in my pursuit of it. Dancy clarifies that I know both of these things: neither is hidden from me and neither idea is a distortion of the other. *Ibid.*, 234.

reason obtains. People sometimes act and want to act to realize certain results because they consider those results to matter independently of their own personal interests and projects.³⁷ It may therefore be undesirable that a full specification of an agent's reason reads:

I have a reason to attend the conference rather than donating this money to famine relief because *I (happen to) have* an interest in someone finding a cure for cancer.

Yet, if we apply this criticism to the definition of agent-relational reasons, the result is far less troubling. If the initial reason statement Dancy describes is read as introducing an agent-*relational* reason, I doubt that it introduces one thought too many into the agent's deliberations because the agent is, after all, aware of the reason being a reason for her to act simply because *she* is in the circumstances that the reason applies to. The fact that she may also think that the reason is also a reason for any rational agent to attend the conference, were they to be in her circumstances, does not undermine this point.

Of course, it may turn out that the agent in question is incorrect in believing that her reason to attend the conference is a reason for every rational agent to act on, were they in her circumstances. That is, she might be mistaken about the reason being an agent-neutral reason. But if this is true, there has to be some perspective, other than the agent's first-personal perspective, from which it can be said that the reason Dancy describes is an agent-*relative* reason if the agent is to be mistaken about it being an agent-neutral reason. (Otherwise, the agent will be right and we have no problem). Consistently, with Norman's case above, such a pronouncement can be made from the

³⁷ Bernard Williams made the contrary point: that it is undesirable for an agent to consider what is required from the moral point of view, before, for example, seeking to rescue her own kin. Williams, 'Persons, Character and Morality', 18.

second-person or third-person perspective. From the third-person perspective, for example, we can say that:

Ralph has a reason to attend the cancer conference (rather than donating this money to famine relief) because Ralph happens to have an interest in someone finding a cure for cancer.

Again, in light of his fourth criticism of Nagel's theory, Dancy may not be entirely satisfied with this response. In reply to Nagel, Dancy argues that we cannot view reasons objectively, as Nagel wants to, because it reduces the agent-relativity of reasons to mere appearances. Dancy is accordingly concerned with insisting on the reality of the reason's form from the first-person perspective. Yet, since, by hypothesis, Ralph *is* in error in thinking that every rational agent in his circumstances has a reason to attend the conference, it is not implausible to claim that the objective significance of her reason is a mere appearance. In any event, in contrast to Nagel, I am not purporting to reduce the content of the agent's reason as viewed from the first-person point of view to the content of her reason as viewed from the third-person standpoint. The reason may be equally real from each viewpoint.³⁸

In light of these considerations we can qualify the definition of agent-relational reasons as only applying from the third-person perspective:

Agent-Relational Justifying Reason*: A justifying reason is *agent-relational* if and only if it contains an essential pronominal or nominal

³⁸ There is one further criticism of Pettit's view that I am aware of, however it does not require extended discussion. David McNaughton and Piers Rawling argue that the reason statement 'all parents have a reason to help their children' has an agent-neutral form, contrary to our judgment that parents act for their children for agent-relative reasons: McNaughton and Rawling, 'Value and Agent-Relative Reasons'. However, Douglas Portmore observes (and McNaughton and Rawling have since accepted) that reasons can also be agent-relative in virtue of referring to a sub-group of the population. My formulation also adds this qualification. Portmore, 'McNaughton and Rawling on the Agent-Relative/Agent-Neutral Distinction'.

back-reference to the agent or a proper sub-group of all actual and possible rational agents when it is fully specified *from the third-person perspective*.

(It may also refer to the agent's circumstances and a certain time, but since relativity to circumstances and time is always present, I leave this aside).

An agent-relational justifying reason will also be agent-relative under the following condition:

Agent-Relative Justifying Reason: If the reason is agent-relational and the reason is not universally possessed by all relevant rational agents then it is also an agent-relative reason. (*e.g. Since Vera does not enjoy driving fast cars, Norman's reason to drive a fast car is not universally possessed*).

Notably, I do not use the word 'possessed' here in a technical sense; I only use it here as a placeholder to identify those agents for whom the justifying reason could be one for which they act. I discuss what is required for this in Chapter 4.³⁹

We can also define an agent-neutral justifying reason:

³⁹ Justifying reasons may also be agent-relative in a sense that pertains to their object domain: those rational agents, if any, for whom the act is being performed. My justifying reason to assist an injured human, *x*, arises for me because *x* is injured, for any human agent who is *x*. In this case the object domain is universal. But if my justifying reason to assist an injured Peta, is thought to arise, not because Peta may be substituted for the variable *x*, but because Peta is an essential rigid designator, the object domain is not universal. John Skorupski describes justifying reasons with a non-universal object domain as not universalizable, rather than as agent-relative. I follow Skorupski in this regard because otherwise my use of the term 'agent-relative' commits me to certain claims in normative ethics by rendering Moral Realism incompatible with a universally possessed reason with a non-universal domain. Such a justifying reason is certainly implausible as a moral reason where justification for it is taken as primitive, as Skorupski notes the irrational egoist takes it. But it may be that non-universalizability is justifiable, not as a primitive, but as appropriate in the circumstances. (Perhaps this is so in fantastical circumstances where only a 'chosen' few could save humanity). But I leave this particular matter aside. Skorupski, 'Agent-Neutrality, Consequentialism, Utilitarianism... A Terminological Note', 50; Skorupski, *The Domain of Reasons*, 69.

Agent-Neutral Justifying Reason: If a justifying reason is agent-relational and universally possessed by members of the appropriate sub-group, then it is an *agent-neutral* reason.⁴⁰

Agent-neutral reasons may be universally possessed *non-contingently* (perhaps, for example, a reason to pursue one's own happiness), or *contingently* (a reason to drive a fast car; a reason to eat the fruit *belle of the night*; a reason to value one's kin—discussed in Chapter 8). It may seem as if there is little point in describing a justifying reason as agent-neutral where this is based on a relatively transient contingency, such as all rational agents having a justifying reason to eat *belle of the night* the morning of a certain date. But given that reasons can be contingently universally possessed in this way, for clarity, I describe justifying reasons that are universally possessed as *agent-neutral* and, where relevant, add that they are *contingently* agent-neutral or *non-contingently* agent-neutral as the case may be.

Finally, the definitions of 'agent-relational', 'agent-relative' and 'agent-neutral' are context-sensitive so that they may apply, not just in regard to all actual and possible rational agents, but also relative to a proper sub-group of that population.⁴¹ This means that it is possible that: *insofar* as one is discussing all actual and possible rational agents, the justifying reason in question is agent-relative. Yet, *insofar* as one is concerned with all actual, rational, human agents, the reason is agent-neutral in virtue of: **1)** the reason being agent-relational; and **2)** all members *of the proper sub-group* possessing the reason. I consider the implications of this in Chapter 3.

⁴⁰ If some agent-neutral reasons have explanatory priority over agent-relational reasons then the definition can be recast as: justifying reasons that entail agent-relational reasons for all members of the appropriate sub-group.

⁴¹ For example, see: Schroeder, *Slaves of the Passions*, 18.

C. Ideal Observer Theory

This account of the distinction between agent-relative and agent-neutral terms is also applicable to propositions, the truth values of propositions, references and facts. I focus on the truth values of propositions, but the application is the same for propositions, references and facts. In this regard, the same idea arises: if the truth value of a proposition is agent-relative, then it is only true (or false) relative to a certain agent, or a certain proper sub-group of the global population of rational agents. If its truth value is agent-neutral, then its truth or falsity is the same for all rational agents. Moreover, where a proposition is false for agent, *A*, its falsity is agent-relational, and either agent-neutral or agent-relative, depending on its truth value for all rational agents of the appropriate sub-group.

The inclusion of ‘agent-neutral dependent relations’ in Moral Impartiality rules in metaethical theories that would not otherwise fall under the scope of Moral Realism. Ideal observer theories do not maintain that moral propositions, truth values, references and facts exist independently of either human agents or their evaluations. Yet, on some versions of the theory, the judgments of an *idealized* observer are argued to be identical for all moral agents. On such a theory the truth of moral propositions is agent-neutral, and in virtue of this and a willingness to interpret moral realism liberally beyond the two guiding ideas, it would be a version of Moral Realism for our purposes.^{42/ 43} Some writers mark this difference by describing such

⁴² See for example: Smith, *The Moral Problem*, 165–6.

⁴³ See n14 in this chapter for writers defining moral realism as consisting of moral facts that are objective or independent in some sense. In contrast to this, Mark van Roojen thinks constructivist and ideal observer theories should be considered as versions of realism and Terence Cuneo regards constructivist theories as one version of paradigmatic moral realism. In addition, Christine Korsgaard describes the theory she defends as ‘procedural realism’, and Michael Smith and Peter Railton categorize their respective ideal observer theories as versions of moral realism: van Roojen, ‘Rationalist Realism and Constructivist Accounts of Morality’, 291–4; Cuneo, *The*

theories as *minimal moral realism* (in contrast to *robust moral realism*).⁴⁴ Other writers refer to such theories as *subjectivism*.⁴⁵ In contrast I simply persist with the definition stated above.

D. Natural Facts

In addition to 'realism' and 'cognitivism', the term 'natural' and its cognates are also terms of art. The most widely employed criterion used to distinguish natural and non-natural facts is one that identifies a natural fact as a fact recognized by scientific disciplines in their ideal state, whereafter the writer provides a list of disciplines she regards as 'scientific'.⁴⁶ Yet, in regard to this type of definition, I agree with Michael Ridge and Nicholas Sturgeon's concern that identifying the disciplines that are 'scientific' requires appealing to some further criterion in virtue of which they are scientific. However, if this is true, then natural facts should be defined by that criterion.⁴⁷ At least 24 different criteria are proposed in the literature to this end, although not all with a view to being endorsed.⁴⁸ In this thesis, I follow Gilbert Harman and

Normative Web, 48; Korsgaard, *The Sources of Normativity*, 36; Smith, *The Moral Problem*; Railton, 'Moral Realism'.

⁴⁴ FitzPatrick, 'Recent Work on Ethical Realism', 749; Joyce, 'Moral Anti-Realism', sec. 1; Brock and Mares, *Realism and Anti-Realism*, 11. Stephen Finlay also makes this distinction using the terms ontological and metaphysical moral realism. Finlay, 'Four Faces of Moral Realism', 821.

⁴⁵ Finlay, 'Four Faces of Moral Realism', 829; Joyce, 'Moral Anti-Realism', sec. 5.

⁴⁶ Adherents to this view include: Moore, 'Preface to the Second Edition', 13; Scanlon, *What We Owe to Each Other*, 57, 58; Scanlon, 'Metaphysics and Morals', 8; Blackburn, *Ruling Passions*, 48; Smith, *The Moral Problem*, 17, 203–4 chap. 2 n1; McNaughton and Rawling, 'Naturalism and Normativity', 29.

⁴⁷ Ridge, 'Moral Non-Naturalism'; Sturgeon, 'Moore on Ethical Naturalism', 555. Also see Michael Huemer who argues that it is the naturalness of the phenomenon that makes the study of it a natural science, not the other way around. Huemer, *Ethical Intuitionism*, 262 n1.

⁴⁸ For example: David Copp defines 'natural facts' as those facts that are known via strong apriority Copp, 'Why Naturalism?', 43; Roger Crisp defines natural facts as those identified by the best scientific theories and conceptual terms available to a being from some non-local point of view: Crisp, 'Naturalism and Non-Naturalism in Ethics', 117; Stephen Finlay defines natural facts as those that are not non-natural facts, where non-natural moral or normative facts are those facts that are

Terence Cuneo's approach in defining natural facts as: those facts identified by a metaethical theory that is predominantly concerned with actually locating moral and morally relevant facts in the world either by the scientific method, or by defining them in a way that allows for their being located this way.⁴⁹ In this regard, both writers acknowledge that this need not mean that the facts identified must perfectly fit the idealized body of scientific knowledge, although as far as possible this is what such a metaethical theory is striving for.⁵⁰ There are, however, disputes in the literature as to whether, *inter alia*, there is a single scientific method, whether it is justifiable and whether it nevertheless remains relativized to sociological, economic and political influences, to a time, a society or the subject matter in question.⁵¹ I can only say that the characterization of the scientific method I follow remains beholden to these matters. With this in mind, I regard the scientific method as consisting in:

formulating a hypothesis about the observations of phenomena, usually in light of a theoretical context and other experiments and observations, and testing predications derived from the hypothesis by conducting replicable experiments from which one may conclude a posteriori that a given hypothesis remains plausible or is proven false.

metaphysically autonomous or *sui generis*: Finlay, 'Four Faces of Moral Realism', 828. For criticism of several possible criteria see: Shafer-Landau, *Moral Realism*, 58–65.

⁴⁹ Harman, 'Is There a Single True Morality?', 29; Cuneo, 'Recent Faces of Moral Nonnaturalism', 854. So explicated, this definition might be understood as a version of what Peter Railton calls methodological naturalism (compared to ontological naturalism). In this regard I should explain that I am only granting the method used to identify natural facts epistemic priority. Natural facts retain explanatory priority: their existence explains our so recognizing them.

⁵⁰ Insofar as metaethicists grapple with the question of explaining how moral facts can exist in the world when we fail to observe them in the same way as other phenomena, this 'external accommodation' method of locating moral facts is sometimes thought to favour Moral Anti-Realism. Of course, Moral Anti-Realism is not however the inevitable outcome.

⁵¹ Barry Gower and Peter Kosso canvas these issues and others, both with a view to ultimately defending the existence of scientific method, in at least some sense. Gower, *Scientific Method*; Kosso, *A Summary of Scientific Method*.

Where a metaethical theory is not predominantly concerned with locating moral facts in the world according to this method or allowing for such, I consider the moral facts it identifies as non-natural. An example of such an approach is one that Cuneo contrasts with the former method. According to this approach, the predominant concern of metaethical inquiry resides in locating moral facts or morally relevant facts for the purpose of constructing a theory 'that comports well with deeply embedded assumptions of ordinary moral thought and practice.'⁵² Among those assumptions, Cuneo lists the following:

that ethical discourse is assertoric in appearance, that moral judgments are intimately linked with appropriate motivation, that entities of various kinds appear to display moral features of various sorts, that some moral obligations appear to govern our behavior regardless of our contingent desires or associations, that we know some moral claims, and so forth.⁵³

Thus, as Stephen Finlay remarks, the aim of this approach is to locate moral facts with a view to vindicating, so far as possible, the appearance of morality or our 'ordinary' understanding of morality (and hence, without regard one way or the other for whether the facts that are identified, if any, fit well with the scientific method).^{54/ 55}

Of course, it is possible that both methods identify one and the same fact as 'moral' and hence the same fact will be a natural moral fact relative to the first approach, and non-natural or perhaps even supernatural relative to the second approach. This result may appear somewhat implausible but I consider it an advantage: it makes sense of the present lack of consensus over criterion purporting to cleanly distinguish natural and non-natural facts. Furthermore, as the terms 'natural' and 'non-natural' are terms of art,

⁵² Cuneo, 'Recent Faces of Moral Nonnaturalism', 854.

⁵³ Id.

⁵⁴ Finlay, 'Four Faces of Moral Realism', 822.

⁵⁵ This approach is sometimes said to favour moral realism in light of, for example, the way we use moral language and our ordinary pretensions of moral objectivity.

what matters most is the aim of the classificatory method employed. Harman and Cuneo's distinction acknowledges this by allowing for approaches with different aims to identify the same fact. So understood then, natural facts are not defined according to a clear cut criterion but instead their identification is constrained by the metaethical aim of actually locating moral facts in the world according to the scientific method in a way that fits as well as possible with idealized scientific doctrine.

II Justifying Reasons

At the commencement of my discussion of agent-neutral and relative reasons I mentioned that I was using these terms in regard to justifying reasons. The literature is rife with discussion of a number of issues about and surrounding justifying reasons as well as the type of reason it is often defined in contrast to, what I shall call *explanatory* reasons. Since Chapters 3-7 in this thesis are concerned with these two types of reasons, it will be worthwhile explaining how I regard this distinction from the outset.

To begin with, the term 'reason' is used in ordinary language in more than one way.⁵⁶ In this thesis I largely set aside its use to describe a faculty or competence of the mind and its use as a point of view or potential source of obligation (as to denote, for example, the principles of practical reasoning, or some kind of authoritative entity, Reason). For the most part I also leave aside its use as a verb ('to reason persuasively you must ...'), its use as a noun, naming the act of reasoning (correctly or incorrectly), its use denoting intuition, and its use denoting a mental state of reasonableness or sensibleness. Rather, I follow some writers in focusing on its use as a count noun, as illustrated in the sentences:

⁵⁶ Most of the taxonomy that follows in this paragraph comes from Mark Schroeder which I need not depart from: Schroeder, *Slaves of the Passions*, 11.

- My having promised to meet Gustave in Rouen is a reason for me to go.
- That Reese saw Augustine running from the house with a revolver is a reason for her to believe that he shot the victim.

These particular illustrations concern a reason to act and believe something respectively and for the most part in this thesis I focus on reasons to act or omit to perform some act. Sometimes the adjectives practical reason and theoretical (or epistemic) reason are used to refer to reasons to act and believe respectively and I briefly discuss the appropriateness of such terms in Chapter 7. For the most part, however, it will suffice to merely understand a practical reason as a reason to act or omit and a theoretical reason as a reason to believe a certain proposition. These two classes of reasons however are not exhaustive. There are also reasons for possessing affective states (for feeling excitement, fear, joy, pride, satisfaction, and sadness) and possibly reasons for having other mental states (for trusting, hoping, wishing, and begrudging).⁵⁷

A. Justifying Reasons

Within the class of reasons to act, writers often, but not always, distinguish between explanatory/ motivating reasons on the one hand, and justifying/ normative reasons on the other.^{58/ 59} Each of these four terms is ambiguous. Explanatory reasons are sometimes understood as reasons that offer an explanation of why an agent was motivated to perform an intentional act, is

⁵⁷ Raz, *Practical Reason and Norms*; Scanlon, *What We Owe to Each Other*; Thomson, 'Normativity'; Skorupski, 'Neutral versus Relative'; Skorupski, *The Domain of Reasons*; Copp, 'Moral Naturalism and Three Grades of Normativity', 256–7.

⁵⁸ Many writers use the terms 'normative' and 'motivating' reasons respectively. See for example: Smith, *The Moral Problem*, 94–6. Allan Gibbard in contrast distinguishes between 'reasons to do' and 'reasons for doing'. Gibbard, 'Reasons Thin and Thick', 290.

⁵⁹ Arguably, this distinction derives from Francis Hutcheson: Hutcheson, 'A System of Moral Philosophy, Volume I: Concerning the Constitution of Human Nature and the Supreme Good', bk. 1.4.III[1]–[5], pp.168–70. Cf. Dancy, *Practical Reality*, 20–25.

motivated to perform an intentional act or would be motivated to perform an intentional act. In contrast, a justifying reason to act is standardly understood as a reason that justifies an intentional *act* (and not: ‘motivation to act’) that an agent has performed, that an agent is performing, or that an agent could perform, because it identifies a consideration that favours what the agent should or should not do.

In the illustrative sentences that follow, the first two sentences exhibit justifying reasons and the latter two exhibit explanatory reasons:

1. That Kira loves her audience is a reason for Kira to play her violin solo well. (Justifying reason)
2. That the Rembrandt is not authentic is a reason for Calvin not to steal it on Friday night. (Justifying reason)
3. That Kira loves her audience is the reason why Kira was motivated to play her violin solo well. (Explanatory reason)
4. That Calvin wants the Rembrandt for his personal collection is the reason why he is motivated to steal it on Friday night. (Explanatory reason)

In one sense, an agent’s explanatory reason for acting may not actually be a justifying reason for her acting. Suppose Miya also wants the Rembrandt for her personal collection and on Saturday night she breaks into the Hermitage Museum to appropriate it. In fact, since Friday night the Rembrandt is in Calvin’s possession, so on Saturday there is no justifying reason for Miya to break into the Hermitage for it. We may, however, be able to cite Miya’s mistaken belief (or something else) as part of an explanatory reason for her conduct. In one sense then, there is an explanatory reason for Miya’s actions,

but no justifying reason and arguably, because of this, the two types of reason are distinct.⁶⁰

However each term, justifying reasons and explanatory reasons, requires disambiguation. I begin with the former. Cases where an agent's explanatory reason for Φ -ing in circumstances C and her justifying reasons for Φ -ing in C diverge have generated a distinction between what are sometimes called subjective and objective justifying reasons, though 'subjective justifying reason' is defined in two slightly different ways. In the first sense, a subjective justifying reason for an agent, A, is a justifying reason that she believes she has given the circumstances as she understands them to be.⁶¹ In the second sense, a subjective justifying reason for A is one that exists for her relative to the circumstances that she is *justified* in believing to exist given her physical and psychological characteristics.⁶² Both senses are contrasted with objective justifying reasons, reasons that A has given the circumstances as they actually are:

⁶⁰ A number of writers maintain that justifying and explanatory reasons are not two types of reasons but are rather two different *senses* of the one type of practical reason. Dancy, *Practical Reality*; Williams, 'Postscript: Some Further Notes on Internal and External Reasons', 93[1]; Schroeder, *Slaves of the Passions*, 14, 14 n21; And in regard to the concept of a reason: Skorupski, *The Domain of Reasons*, 53; cf. Parfit, *Reasons and Persons*, 118; Pettit and Smith, 'External Reasons'. These writers maintain that there is only one type of reason (in regard to this point) because justifying reasons and explanatory reasons are identical or otherwise closely related. I actually think that there is a necessary relation between the two and I argue for this in Chapters 4-6. This, in itself, does not entail (nor do writers consider that it entails) that the two reasons are not of different types. The point requires defence, which some writers provide. Hereafter, I assume that the two reasons are distinct types of reasons, although nothing depends on this.

⁶¹ See for example: Lenman, 'Reasons for Action: Justification vs. Explanation', sec. 1. Roger Crisp also distinguishes this sense of subjective justifying reasons from objective justifying reasons, albeit in slightly different terms: Crisp, *Reasons and the Good*, 39.

⁶² Richard Joyce provides a version of the distinction between this sense of subjective justifying reasons and objective justifying reasons. Joyce, *The Myth of Morality*, 53. Dancy also draws this distinction in other terms. Dancy, *Practical Reality*, 107.

Subjective Justifying Reason (1): a reason for an agent to Φ (perform some act) that she believes she has given the circumstances as she understands them to be.

Subjective Justifying Reason (2): a reason for an agent to Φ relative to the circumstances that she is *justified in believing* to exist given her physical and psychological characteristics.

Objective Justifying Reason: a reason for an agent to Φ given the circumstances as they are in fact.

Reference to subjective justifying reasons is motivated by two considerations. Firstly, as we have seen there is sometimes the need to explain an agent's conduct by referring to the justifying reason that the agent erroneously believed she had to act. Secondly, there is at times the need to not only explain the agent's conduct but also justify it, given the epistemic and computational limitations of rational human agents. For instance, although the old sunken ship is not at Site A, the recovery team has significant evidence to believe that it is and so they have a subjective justifying reason (in both sense (1) and (2)) to send divers down at Site A.

Sometimes writers avoid using the term 'justifying reason' in light of a concern for its potential confusion with subjective justifying reasons.⁶³ They accordingly use the term 'normative reason' instead. Yet whichever term is used, one still needs to distinguish between subjective and objective justifying reasons. I retain the term 'justifying reason' because I find the adjectival term 'normative' to be especially vague. Unless indicated otherwise, by 'justifying reason' I mean 'objective justifying reason'.

In regard to both subjective and objective justifying reasons writers standardly distinguish between *pro tanto* and *conclusive* (or 'all-things-considered') justifying reasons. Whereas a *pro tanto* reason provides a justifying reason in favour or disfavour of performing a certain act, I shall

⁶³ For example: Dancy, *Practical Reality*, 6–7; Crisp, *Reasons and the Good*, 39.

understand a conclusive reason as not adding to the weight of the pro tanto reason(s) that favours performing that act. Rather, as Jonathan Dancy uses the term, a conclusive reason is merely an expression of the pro tanto reason(s) that is the most influential.⁶⁴ Given this, when discussing a conclusive reason I am only discussing it in light of the pro tanto reason(s) that supports it, and hence, not in terms of the *totality* of pro tanto reasons (favourable and unfavourable) that may have influenced an agent in coming to the conclusive reason she comes to.⁶⁵ Furthermore, where a pro tanto justifying to perform a certain act (Φ) is, in itself, (or in combination with other pro tanto justifying reasons) of sufficient weight to justify the relevant agent in Φ -ing for that reason, we can say that it is (or that there is) a *sufficient* pro tanto justifying reason for the agent to Φ . Where it is not of sufficient weight, we can say that it is an *insufficient* pro tanto justifying reason for the agent to Φ .⁶⁶ A rational agent may have more than one sufficient reason to act. After walking in the forest, Hailey may have a sufficient subjective reason to turn left and a sufficient subjective reason to instead turn right, perhaps because she is lost. Alternatively, she may have a sufficient subjective (and objective) reason to turn left and a sufficient subjective (and objective) reason to turn right because both paths lead home. However, as she correctly believes that the path to her left is slightly shorter, then, depending on other details, her sufficient reason to turn left may be of slightly greater weight than her sufficient reason to turn right.

⁶⁴ Dancy, 'What Do Reasons Do?', 39–40.

⁶⁵ A similar distinction arises for moral *principles*. See: Dancy, 'Moral Particularism', sec. 1.

⁶⁶ See also John Skorupski's distinction between pro tanto, sufficient and overall reasons to act and his argument for their irreducibility: Skorupski, 'The Unity and Diversity of Reasons'; Skorupski, *The Domain of Reasons*, 37–41.

B. Explanatory Reasons

The term ‘explanatory reason’ also requires disambiguation firstly because it is used in senses in which it takes different objects. We defined the term explanatory reason above in regard to actions as a reason that offers an explanation of why an agent was *motivated* to perform an intentional act, is motivated to perform an intentional act or would be motivated to perform an intentional act. This definition must be distinguished from another that is commonly employed: a reason that explains an intentional *act* (compared to motivation for acting).⁶⁷ The two types of explanatory reason are often conflated, perhaps due to the ambiguity Alfred Mele observes in the term ‘motivation’ which admits of a ‘success’ reading and an instantiation reading.⁶⁸ On the success reading the agent’s motivation results in the agent performing the relevant intentional act. In contrast, the instantiation reading is silent on the question of whether the agent’s motivation translates into action. Unless otherwise noted, I adopt the instantiation reading: an ‘explanatory reason’ explains an agent’s *motivation* to act, however little the motivation may be and whether or not it results in action.

Explanatory reasons also admit other objects. For example, they provide causal explanations of unintended conduct, such as tripping over one’s feet, and causal explanations of natural events, such as rainstorms. Explanations of this kind may occur at more than one level. For example, it may be that Jack’s tripping over his feet can be causally explained by his psychological and affective states (his beliefs and resentment, for example) or by his neurochemistry, character traits or by facts in the world itself. I discuss this further in Chapter 3.

⁶⁷ James Lenman discusses this distinction in more detail: Lenman, ‘Reasons for Action: Justification vs. Explanation’.

⁶⁸ Mele, *Motivation and Agency*, 14.

The term is also ambiguous in virtue of there being more than one consideration that can explain an agent's motivation. In this regard, it is somewhat common for writers to distinguish between: **a)** a causal or teleological explanation of an agent's conduct or motivation; and **b)** an agent's (objective or subjective) *justifying* reason to intentionally act which explains why the action is worth doing.⁶⁹ But we can also distinguish between an *agent's own* explanation of her conduct (a subjective explanatory reason) and the actual explanation of her conduct (an objective explanatory reason). For example, if asked why he stole the Rembrandt Calvin may say that he wanted it for his personal collection. But this might be false: it may be that Calvin actually stole the painting because Kirk doubted that he could do it.

Fortunately, for the most part in this thesis I only use the term 'explanatory reason' to mean an *objective* explanatory reason for an agent's motivation to intentionally act and I expressly make the distinction between subjective explanatory reasons where relevant. At this point, I remain neutral on the question of whether this reason is best given by a causal explanation (at a certain level of detail) or by the agent's objective or subjective justifying reason.

III Outlook

With the above distinctions in hand, Chapter 2 considers versions of G. E. Moore's Open-Question Argument against Moral Realism. Moore's own

⁶⁹ Dancy, *Practical Reality*, 7. The following writers make a similar point in regard to the term 'motivating reason'. It can be used to denote a causal or teleological explanation given from the third person point of view and it can be used to denote the subjective justifying reason that motivated the agent to act: Finlay, 'The Reasons That Matter', 16; Korsgaard, 'Acting for a Reason', 208–9; Raz, 'Reasons: Normative and Explanatory', 195; Schroeder, *Slaves of the Passions*, 12–4; I also understand Roger Crisp's distinction in this way: Crisp, *Reasons and the Good*, 38.

version of the argument fell into relative disfavour in the last quarter of the twentieth century for reasons which are not difficult to come by. Yet, significantly less noted is the number of contemporary writers who persist in defending a version of the argument against objections that are still rather perfunctorily recited against it as well as the argument's relation to work in the philosophy of mind.

Chapter 3 turns to consider a more recent argument against Moral Realism, what I call The Relative Reasons Argument, which relies on a relation between moral facts and reasons for agents to act. Specifically, this chapter considers a premise in this argument: a variant of The Humean Theory of Motivation. While it is concluded that this premise is false, proof of the truth of the two remaining premises in The Relative Reasons Argument nevertheless establishes a certain dependency of moral facts on moral agents' motivational capacities that many contemporary authors deny and would claim that, if true, it would disprove moral realism.

As we shall see, however, both premises are defensible and Chapters 4–6 defend a different relation between agents' justifying reasons and their recognitional and motivational capacities than that often cited. Chapter 7 then contends that a particular type of relation exists between moral facts and agents' justifying reasons. Together, the two relations establish a relation between moral facts and agents' recognitional and motivational capacities.

Chapter 8 considers the implications of this overall relation and denies the view that moral facts exist in a certain robust or objective sense while nevertheless being dependent on agents' recognitional and motivational capacities. Instead, it is maintained that moral agents play a much larger role in the determination of moral facts than many moral realists acknowledge. Specifically, I defend the plausibility of a conservative ideal observer theory that I call Judicious Agent Theory.

Moreover, while Judicious Agent Theory is a version of Moral Realism it is not a naturalist theory: the process of testing predications from hypotheses by replicable experiments does not, in itself, do enough work to permit one to identify moral facts. Independent theoretical work is required. Ultimately though, I am less interested in how Judicious Agent Theory is characterized than in what is being characterized.

2

Open Questions & Contemporary Answers

‘Yes,’ echoed Dorian, leaning back in his chair, and looking at Lord Henry over the heavy clusters of purple-lipped irises that stood in the centre of the table, ‘what do you mean by good, Harry?’

‘To be good is to be in harmony with one’s self,’ he replied, touching the thin stem of his glass with his pale, fine-pointed fingers. ‘Discord is to be forced to be in harmony with others. One’s own life — that is the important thing. As for the lives of one’s neighbours, if one wishes to be a prig or a Puritan, one can flaunt one’s moral views about them, but they are not one’s concern. ...

‘But, surely, if one lives merely for one’s self, Harry, one pays a terrible price for doing so?’ suggested the painter.

‘Yes, we are overcharged for everything nowadays.

——Oscar Wilde, *The Picture of Dorian Gray*, chap. VI

Chapter 1 defined Natural Moral Realism as maintaining that some moral propositions are true in virtue of correctly representing certain natural facts. It remains for the proponent of a version of this view to identify which natural facts these are. An initially influential objection from G. E. Moore promises that the pursuit of such an endeavour is in vain. Moore considered the predicate ‘morally good’ to be uniquely primitive in the sense that it cannot be defined by terms within or outside of moral discourse. While Moore may or may not be in error in claiming that moral goodness cannot be defined in virtue of other moral or normative terms, his argument is equally applicable to other moral predicates so I leave aside the issue of which moral or normative predicate(s) are primitive within moral discourse.

Moore's Open-Question Argument ('OQA') purported to establish that the predicate 'morally good' (hereafter 'G') is indefinable.¹ There is lack of agreement as to whether Moore's use of this predicate, G, was intended to denote, *inter alia*, the sense of the predicate 'moral good', or both its sense and at other times its referent. On the one hand, Moore clearly expresses his intent in *Principia Ethica* ('PE') to go beyond semantics to investigate properties.^{2/ 3} However, on the other hand, Moore's arguments for the indefinability of G are limited to proving that *the predicate* G is indefinable. Given Moore's express reference to G as a property the most plausible interpretation of his relaxed transition from discussion of G on the semantic level to discussion of G on the ontological level is that he *was* concerned with ontology and simply assumed in *PE* that a distinct moral predicate solely refers to its own distinct property.⁴ In light of this, I expressly preserve this

¹ In his unpublished and incomplete 'Preface to the Second Edition', (hereafter 'the Second Preface') Moore states that he was concerned for the predicate 'good' in 'that sense of the word "good," which has to do with the conceptions of "right" and "wrong," a relation, which makes it *the* sense which is of the most fundamental importance for Ethics. Let us call the predicate G.' Moore, 'Preface to the Second Edition', 5 (emphasis present).

² Moore, *Principia Ethica*, 54, §2: 'for verbal questions are properly left to the writers of dictionaries [...] philosophy, as we shall see, has no concern with them'; 58, §6: 'Such a definition can never be of ultimate importance in any study except lexicography.'; 60, §8: 'But (3) we may, when we define horse, mean something much more important. We may mean that a certain object, which we all of us know, is composed in a certain manner: that it has four legs, a head, a heart, a liver, etc., [...] It is in this sense that I deny good to be definable.' In §9, in contrast, Moore expressly refers to 'good' as an adjective when he distinguishes it from the substantives which possess it, 'the good'. However, in the opening sentence of §10, Moore iterates that the sense of 'definition' he is concerned with is ontological and he proceeds in this section to refer to G as a 'quality' and then at the end of §10, as a property. The Second Preface similarly suffers with these ambiguities.

³ For theories in favour of the view that Moore's thesis enters the ontological plane see: Pigden, 'Desiring to Desire: Russell, Lewis, and G. E. Moore', 251–2; Fumerton, 'Open Questions and the Nature of Philosophical Analysis', 233; Nuccetelli and Seay, 'What's Right with the Open Question Argument', 264 n4; Baldwin, 'Logic and Philosophical Analysis', 418; Schroeder, 'Realism and Reduction', 4.

⁴ Moore may have been influenced by the Socratic view of using language as a mirror of reality: Socrates, 'Phaedo', p.240, 99e–100a. Alternatively, Moore's assumption here may have been influenced by his departure from British Idealism. Hilary Putnam and Richard Fumerton, for example, argue that Moore conflated

distinction by letting 'G' denote the *predicate* 'morally good' and by letting 'G_R' denote G's referent.

Since its publication, there has been significant criticism of Moore's OQA, including by Moore himself.⁵ The criticism is such that in discussing OQA Nicholas Sturgeon recently remarks 'I will try not to linger beating dead horses'.⁶ I largely agree with Sturgeon, however, firstly, more than one modern writer defends a close reconstruction of the argument.⁷ Secondly, some writers, while not defending it outright, have pointed replies to certain objections that have been influential against OQA.⁸ Thirdly, many writers who think that Moore's argument and close reconstructions of it are unsound have nevertheless asserted that it does point to something that

concepts and properties and Putnam explains in an earlier context, not referring to Moore, that an older notion of 'property' denotes 'predicate' (cf. the modern notion of 'physical property') and that the older notion was usually interchangeable with 'concept'. Putnam, *Reason, Truth and History*, 84, 207; Fumerton, 'Open Questions and the Nature of Philosophical Analysis', 233. For an argument that Moore did not conflate concepts and properties, whether PE was concerned with ontology or not see: Stratton-Lake, 'Introduction', 9.

⁵ Moore, 'G.E. Moore: A Reply to My Critics'. Although, writers have mentioned that Moore was somewhat self-effacing.

⁶ Sturgeon, 'Moore on Ethical Naturalism', 532.

⁷ Robert Adams denies that OQA succeeds in its original context and against non-evaluative analyses, but defends a version of it against Richard Boyd's definition of value terms and in defence of his own supernatural definition: Adams, *Finite and Infinite Goods*, 77–82. Michael Huemer argues that OQA succeeds against analytic reductionism: Huemer, *Ethical Intuitionism*, 68–71. Ralph Wedgwood employs Christine Korsgaard's normative question to the 'substantive recognitional view' and the 'constructivist view' of reasons in the same way as OQA: Wedgwood, 'Choosing Rationally and Choosing Correctly'. For reasons similar to those upon which OQA is based in regard to 'good', Allan Gibbard thinks that one cannot give the meaning of 'rational' in terms of principles. Gibbard, *Wise Choices, Apt Feelings*, 11. Stephen W. Ball uses OQA as a linguistic test. Ball, 'Linguistic Intuitions and Varieties of Ethical Naturalism'. Chris Heathwood defends A. J. Ayer's formulation of OQA, arguing that it relies on a logico-linguistic intuition for evidence: Heathwood, 'Moral and Epistemic Open-Question Arguments', 88. Susana Nuccetelli and Gary Seay recently defend semantic and metaphysical reconstructions of OQA: Nuccetelli and Seay, 'What's Right with the Open Question Argument'. Finally, John Skorupski uses the argument in regard to 'goodness' and 'reason relations' and, more generally, all normative predicates and properties: Skorupski, *The Domain of Reasons*, 447–51.

⁸ For example: Smith, 'Moral Realism'.

could forcefully underwrite Moore's concern.⁹ Finally, the same pattern of objections to OQA arises in response to The Knowledge Argument in philosophy of the mind.

In this chapter I consider contemporary replies to significant objections levelled against Moore's OQA, in conjunction with replies to analogous objections to The Knowledge Argument. I contend that contemporary replies to two of these four objections are ultimately unsuccessful and that Moore's OQA therefore leaves Natural Moral Realism untroubled. It does however serve as the foundation for a more sophisticated argument to which we turn in Chapters 3-7.

I The Open-Question Argument

The macro-structure of Moore's OQA in *PE*, §13, is eliminative in nature: Moore firstly denies that G_R could be a complex property, he then denies that G_R could mean 'nothing at all', and concludes that G_R must be a simple property. Some writers use the singular term 'OQA' to refer to the entire eliminative argument;¹⁰ others employ this term in reference to the first sub-argument.¹¹ To maintain the distinction between the two sub-arguments I use the name 'OQA (Complex Property)' to denote Moore's first sub-argument, and 'OQA (No Property)' to denote his second sub-argument. Writers interpret Moore's OQA (Complex Property) in different ways and while my concern does not lie in correctly interpreting Moore's argument I note alternative interpretations where appropriate.

⁹ Darwall, Gibbard, and Railton, 'Toward *Fin de siècle* Ethics: Some Trends', 116; Blackburn, *Ruling Passions*, 86; Sayre-McCord, 'Moral Realism [SEP]', sec. 3.

¹⁰ For example: Baldwin, *G.E. Moore*, 87–8.

¹¹ Bruening, *The Is-ought Problem*, 13; Shafer-Landau, *Moral Realism*, 57; Darwall, 'Moore, Normativity, and Intrinsic Value', 481; Daly, *An Introduction to Philosophical Methods*, 52.

OQA (Complex Property) is largely captured in the opening sentence of §13(1):

The hypothesis that disagreement about the meaning of good is disagreement with regard to the correct analysis of a given whole, may be most plainly seen to be incorrect by consideration of the fact that, whatever definition be offered, it may be always asked, with significance, of the complex so defined, whether it is itself good.¹²

In the remainder of the passage Moore supports this contention—that any definition of G will be incorrect—with one illustration. The language Moore uses thereafter is both inconsistent and ambiguous,¹³ however common to his assertions, and most interpretations, is a request to compare two questions with the supposed difference between the two demonstrating that G cannot be defined as intended. I reconstruct Moore's OQA (Complex Property) as follows:

OQA (Complex Property): Modus Tollens

- (1) If the predicate G has the same internal meaning¹⁴ as 'that which we desire to desire' ('D'), then whenever, and only whenever, the

¹² Moore, *Principia Ethica*, 67, §13.

¹³ For further on these difficulties see: Feldman, 'The Open Question Argument: What It Isn't; and What It Is', sec. 3 where he acknowledges problems arising from: **a)** Moore's ignorance of the use/ mention distinction; **b)** his blurring of the distinction between the use of a term as a predicate and the use of a term as the name and noun of the property expressed by the predicate; **c)** his ambiguous use of the verb 'to be' and; **d)** his inconsistent use of linguistic terminology. In the Second Preface Moore himself acknowledges that he sometimes speaks as if the naturalistic fallacy is the confusion of the 'is' of predication and the 'is' of identity or between the 'is' of identity and the 'is' of including one class in another. This would also explain his confusion in OQA (Complex Property): Moore, 'Preface to the Second Edition', 20.

¹⁴ I use the singular term 'internal meaning' instead of merely 'meaning' to distinguish the sense of meaning I wish to employ from the sense employed in external theories of semantics. External theories of semantics assert that the meaning of a singular or general term is identical to or lies in the ontological entity that the term refers to. If the meaning of the predicates in question is understood according to this latter theory, OQA (Complex Property) is open to challenge, as Brink argues it is, because two predicates may have different connotations for the speaker and yet have identical meanings. Brink, 'Realism, Naturalism and Moral Semantics', 158–61.

first predicate is truly predicated of a given subject, the second predicate will also be truly predicated of that subject.

- (2) Whenever either G or D is truly predicated of a given subject it is not always the case that the other predicate is also truly predicated of that subject.

Sub-Argument for premise (2): We have a linguistic intuition favouring premise (2) that is generated in two steps:

- (i) Suppose we accept that G has the same internal meaning as D.
- (ii) Nonetheless, for us, Question A below does not trivially yield an affirmative answer in the same way that Question B does:

Question A: Can a subject, χ , that is D, fail to be G?

Question B: Can a subject, χ , that is D, fail to be D?

(C₁) G does not have the same internal meaning as D.

Moore then argues for the generalization of C₁.

OQA (Complex Property): Generalization

(C₁)

- (3) The proposal that any other predicate has the same internal meaning as G can be successfully denied in the same way as in OQA (Complex Property): Modus Tollens.

(C₂) G does not have the same internal meaning as any other predicate.¹⁵

¹⁵ Henry Sidgwick published an argument in similar terms before PE, although in the context Sidgwick was writing in I think it is at least arguable that the scope of his argument is significantly restricted compared to Moore's: Sidgwick, *Methods of Ethics*, bk. I, chap. IX, §2, p.109. Moreover, Sidgwick may have later withdrawn this argument. Hurka, 'Moore in the Middle', 602.

Interpretations of Moore's OQA (Complex Property) vary from the argument I state here along more than one dimension and in this regard two comments are in order. Firstly, I interpret the argument as only relevant for competent speakers of the language within which it is presented. Secondly, the meaning of 'trivially' requires clarification. In contrast to trivially yielding a certain answer Moore himself says that Question A has 'significance',¹⁶ is 'intelligible',¹⁷ 'synthetic'¹⁸ and in a different context, that it is an 'open question'.¹⁹ Furthermore, in the Second Preface he clarifies that by 'synthetic' in *PE*, he meant 'non-tautologous', though he then adds further clarifications which cloud the matter²⁰ and writers disagree on the criterion that must be met for a question to be significant or open.²¹ In OQA (Complex Property) and OQA (No Property) I use the word 'trivially' in the following way:

- **Trivially yields:** a question trivially yields an answer iff the answer to the question is knowable *per se notum*, (that is, knowable in virtue of understanding the meaning of the terms themselves).²²

¹⁶ For example: 'it may be always asked, with significance, of the complex so defined, whether it is itself good.' Moore, *Principia Ethica*, 67, §13.

¹⁷ See: 'it is apparent, on a little reflection, that this question is itself as intelligible, as the original question "Is A good?"' id.

¹⁸ 'That propositions about the good are all of them synthetic and never analytic; and that is plainly no trivial matter.' *ibid.*, 58, §6.

¹⁹ *Ibid.*, 72, §14, 95, §27. (According to the OED, use of this term arose in English in 1532).

²⁰ Moore, 'Preface to the Second Edition', 10, 22. After reflecting on the matter, Moore explains that he should have said that 'G is not identical with any predicate of this particular class, [...]' where 'class' refers to natural and metaphysical predicates or predicates designating either contingent or intrinsic properties. *Ibid.*, 11.

²¹ Hancock, 'The Refutation of Naturalism in Moore and Hare', 326–34; Snare, 'The Open Question as a Linguistic Test'; Nuccetelli and Seay, 'What's Right with the Open Question Argument'; Ball, 'Reductionism in Ethics and Science: A Contemporary Look at G.E. Moore's Open-Question Argument', 207; Clark and Clark, 'What Is Goodness?', 154.

²² I am reluctant to use the term 'tautology' here because, in one sense, a word or phrase may be a tautology merely because the same word or phrase is repeated. I also eschew using the term 'analytic' given its various uses in philosophical history. If pressed to distinguish the sense of 'analytic' I intend here it would be the sense in

- **Does not trivially yield:** a question does not trivially yield an answer iff the answer to the question is not knowable *per se notum*.

Significantly then, the difference between Questions A and B is to be determined according to the meaning of each question. It is not to be determined according to variations in the sentence-type.²³ So clarified, OQA (Complex Property) exacts a test of synonymy, a test notoriously problematic due to the difficulties of stating precise criteria for it. The idea of OQA (Complex Property) then is to demonstrate that, even when one thinks that a purported definition of G is correct, one's doubt in affirming the truth of a question in the form of Question A serves to show that the definition is unsuccessful. Accordingly, G possesses an internal meaning that differs to any purported definition of it.^{24 / 25}

which I understand Leibniz, Kant (in one sense of his ambiguous use of the term) and John Stuart Mill to mean by it: to denote that the predicate is already contained in the subject, at least implicitly, as in: 'All ailurophobes are afraid of cats'. But again, this definition does not quite make the point that in virtue of knowing the meaning of the predicate in question, one knows that the assertion is true.

²³ Richard Fumerton and Chris Daly make this distinction: Fumerton, 'Open Questions and the Nature of Philosophical Analysis'; Daly, *An Introduction to Philosophical Methods*, 54–6. An interpretation of 'trivially' determined according to variation in sentence-type would mean that, even if a purported definition of G were true, OQA (Complex Property) would necessarily still be sound because Questions A and B will still include different predicates (whether or not the overall meaning of each question is the same).

²⁴ Alternative questions with the same type of relation can generate the same 'open' phenomenon. For one example: 'Can A (a subject predicated with G) be predicated with J?'; 'Can A (a subject predicted with G) be predicated with G?' Horgan and Timmons note this first alternative: Horgan and Timmons, 'New Wave Moral Realism Meets Moral Twin Earth', 461.

²⁵ The interpretation of Questions A and B in premise (1) varies among writers. For example, Stratton-Lake and Hooker distinguish between an OQA in regard to *properties* and an OQA in regard to *objects*. 'According to the property version, the open question is "Is it good that A is *n*?", where '*n*' is the natural property with which good is to be identified.' According to the object version, the question is 'Is an A that is *n* good?' Stratton-Lake and Hooker, 'Scanlon versus Moore on Goodness', 150–1 (emphasis present). Furthermore, while the two questions as I present them include predication relations, it is also argued that they comprise: self-predication relations (Altman, 'Breathing Life into a Dead Argument: G.E. Moore and the Open Question Argument', 395–408); and relations that vary according to which of the supposedly multiple open-question arguments in §13 one is focussing on (Feldman, 'The Open Question Argument: What It Isn't; and What It Is', 31ff). Feldman does,

Whereas OQA (Complex Property) is designed to prove that G does not refer to a complex property, OQA (No Property) is designed to show that G cannot mean nothing at all, and hence to leave Moore with the conclusion that G is a simple notion.²⁶ However, all parties to the debate agree that G has a meaning. Philosophers disagree on *what* its meaning but this is a different matter. As Paul Bloomfield remarks: 'Even if expressivism is right and 'good' can no more be defined than 'hooray', we would not thereby conclude that 'good' is meaningless.'²⁷

Moore may have thought that proving this conclusion is important because of his assumption that a distinct predicate solely refers to its own property and because he was then keen to show that there *is* such a property. For this reason, if any concern for Moore's argument that 'G does not mean nothing at all' were to arise it would arise in regard to his assumption that only G refers to G_R (and hence, this argument would relevantly show that G does not refer to 'nothing at all').²⁸ The argument implicit in Moore's assumption is explicated in OQA (Referent) and begins with (C₂) from OQA (Complex Property):

however, argue that each test in §13 expressly or implicitly relies on the principle of compositionality, according to which, the meaning of a whole statement is determined by the meanings of its parts. On this principle, G will only be successfully defined as a term which is synonymous with G, and a term that accordingly preserves G's meaning after it is substituted for G.

²⁶ Moore, *Principia Ethica*, 68–9, §13.

²⁷ Bloomfield, 'Opening Questions, Following Rules', 176. I may add here that since, on the contrary, theories of expressivism (cf. emotivism) often do not merely describe one's state of mind or one's feelings but rather voice or express a view with the purpose of gaining other people's consensus, expressivism considers the use of moral terms to be meaningful in light of this end.

²⁸ In addition to this interpretation, it may also be thought that this passage provides indirect support for premise (3), OQA (Complex Property) by illustrating that G cannot be defined as 'pleasure' and as indirect support for interpretations of the relevant two questions Moore wants to ask in premise (1), OQA (Complex Property). See, for example: Baldwin, *G.E. Moore*, chap. 3, sec. 5.

OQA (Referent)

- (C₂) G does not have the same internal meaning as any other predicate.
- (4) The meaning of a predicate is the referent for which it stands. *(Thus, if one predicate (A) has a different meaning to another predicate (B), then predicate A will refer to a referent that is different to the referent that predicate B refers to).*

-
- (C₃) G_R is not identical to any other referent; any purported identification of G_R can always be successfully denied.

II The Knowledge Argument

The structure of OQA (Complex Property) and OQA (Referent) is not unique. Firstly compare:

According to the pragmatists, to say 'it is true that other people exist' *means* 'it is useful to believe that other people exist'. But if so, then these two phrases are merely different words for the same proposition; therefore when I believe the one I believe the other. If this were so, there could be no transition from the one to the other, as plainly there is. This shows that the word 'true' represents for us a different idea from that represented by the phrase 'useful to believe' [...] ²⁹

Similarly, Frank Jackson's Knowledge Argument consists in a thought experiment designed to show that theories of physicalism in the philosophy of the mind to that date have not explained phenomenal experiences: ³⁰

Mary is confined to a black-and-white room, is educated through black-and-white books and through lectures relayed on black-and-white television. In this way she learns everything there is to know about the physical nature of

²⁹ Russell, 'William James's Conception of Truth', 112 (emphasis present). Cf.: James, 'Two English Critics', chap. 14.

³⁰ Notably, Jackson changed his mind in 1995 about physicalism and responds to The Knowledge Argument by arguing that although we cannot 'place' phenomenal concepts (and perhaps moral concepts and our concept of free will) we have to develop a replacement concept. Jackson, 'Postscript'.

the world. She knows all the physical facts about us and our environment, in a wide sense of 'physical' which includes everything in *completed* physics, chemistry and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this, including of course functional roles. If physicalism is true, she knows all there is to know. For to suppose otherwise is to suppose that there is more to know than every physical fact, and that is just what physicalism denies. [...] It seems, however, that Mary does not know all there is to know. For when she is let out of the black-and-white room or given a color television, she will learn what it is like to see something red, say. This is rightly described learning—she will not say “ho, hum.” Hence, physicalism is false.³¹

In the same paper, Jackson clarifies that the relevant intuition invoked is one about Mary's *knowledge* and not one about her imagination. If physicalism is true, he claims, Mary *would* know what it is like to sense red and she would not therefore need to rely on imagining this. Thus, despite Mary's brilliant understanding of neurophysiology and the sciences, the intuition purportedly elicited is that she would not know what it is like to sense red; it is not that she could not imagine this.³²

Jackson also clarifies that the relevant intuition is in regard to Mary's lack of knowledge about the experiences of *others*, not about her own experiences. Thus, everyone can agree that Mary will learn something after being released merely in virtue of having new experiences. However after she sees her first ripe tomato Jackson explains that she will realize how impoverished her conception of the mental life of *others* has been while she was in the black-and-white room, despite all her learning there of the physical facts of the neurophysiologies of others.

Jackson states his most recent version of the argument, in the form of *modus tollens*:^{33 / 34}

³¹ Jackson, 'What Mary Didn't Know', 51 (emphasis present).

³² *Ibid.*, 52.

³³ Jackson, 'Foreword: Looking Back on the Knowledge Argument', xvii (emphasis added).

³⁴ In the paper that David Chalmers distinguishes 'hard' and 'easy' problems of

The Knowledge Argument

- I. If physicalism is true, then complete physical knowledge is complete knowledge *tout court*.
- II. Complete physical knowledge is *not* complete knowledge *tout court* (at least not as far as the mind is concerned).

Sub-Argument for premise II: We have an intuition in favour of premise II developed in two steps:

- (i) Mary has complete physical knowledge before leaving the black-and-white room.
- (ii) Mary would learn something about the colour experience of others upon leaving the black-and-white room.

Conclusion: Physicalism is not true.

We can see that both OQA (Complex Property): Modus Tollens and The Knowledge Argument proceed in modus tollens, seek to deny the consequent by eliciting an intuition in favour of a particular counterexample, and conclude that ontological entities of a certain kind are not reducible to natural and physical facts respectively. Moreover, just as Moore's OQA (Complex Property) purports to strike down all theories of natural moral realism, The Knowledge Argument is equipped with the same purport in regard to physicalism.

For all this, The Knowledge Argument varies from OQA (Complex Property) in two significant ways. Firstly, The Knowledge Argument does not make the argument for the generalization to other colours, qualia, and phenomena explicit in the way it is made explicit in OQA (Complex Property): Generalization. This is partly a result of the way Jackson formulates The Knowledge Argument but it is also because the intuition that colours in

consciousness, Chalmers' definition of the hard problem bears a striking resemblance to OQA-like arguments. Chalmers, 'Facing Up to the Problem of Consciousness', secs. 3, 5.

general are not reducible to physical facts is easier to glean from the thought experiment about Mary than the linguistic test in Sub-Argument for premise (2) in OQA (Complex Property). The conclusion of OQA (Complex Property) denies that all moral predicates can be defined in singular and general natural terms by showing that G is indefinable, just as the conclusion of The Knowledge Argument denies that all facts can be reduced to physical facts by showing that a fact about others' experience of red is irreducible.

The second point distinguishing OQA (Complex Property) and The Knowledge Argument is that the former argument relies on a linguistic intuition to purportedly deny that moral terms could have a meaning expressible in entirely natural terms. The Knowledge Argument in contrast relies on an epistemic intuition to deny that all facts are reducible to physical facts. As we shall see this difference is significant even though the class of replies to each argument are compellingly similar.

III Objection 1: Vicious Circularity

A. OQA

The first objection takes issue with the truth of the second step in Sub-Argument for premise (2), OQA (Complex Property) and, as a result, also the generalization of C₁ in premise (3).³⁵ William Frankena firstly maintains that Moore's naturalistic fallacy is the result of the more general error ('the definist fallacy') of identifying one property with another property when in fact the two are not identical. Secondly, Frankena maintains that when arguing that the identification of a natural property with G_R is the result of the definist error, Moore fails to justify his assertion that there *are* two

³⁵ Frankena, 'The Naturalistic Fallacy'.

distinct properties under observation (and not one) and so in merely assuming this, Moore begs the question against the possibility of such identification.³⁶

Frankena's objection accordingly disputes the truth of the second step of Sub-Argument for premise (2) on the basis that Moore's implicit justification for its truth derives from a presupposed belief that G is indefinable, a belief that is the conclusion he purports to establish. Thus, if G really *did* mean D as step (i) claims, then both questions asked in step (ii) of Sub-Argument for premise (2) *would* trivially yield an affirmative answer and one could say: 'One cannot *really* ask with significance if G means D.' Frankena concludes that whether or not G means D accordingly requires deciding 'by whatever method we may find satisfactory for determining whether or not a word stands for a characteristic at all'³⁷ and that choice of method in this regard is a problem arising in other philosophical areas.³⁸

In reply, Stephen W. Ball argues that Moore's OQA (Complex Property) invokes pre-philosophical linguistic intuitions of native speakers who need not have any linguistic expertise.³⁹ Ball explains that Moore's OQA (Complex Property) is addressed to a philosophical audience who are philosophically informed, in a sense, but the data that OQA (Complex Property) relies on does not presuppose such knowledge. Accordingly, Ball's point is that in intuitively judging that Question A in Sub-Argument for premise (2) does not trivially yield an affirmative answer, Moore is not presupposing that D cannot be defined as G, rather, Moore is (or may be) merely noticing that pre-philosophic intuitions provide evidence that D is incorrect as a definition for G.

³⁶ Ibid., 473.

³⁷ Frankena, 'The Naturalistic Fallacy', 477.

³⁸ For a variation on Frankena's argument see: Hancock, 'The Refutation of Naturalism in Moore and Hare'.

³⁹ Ball, 'Linguistic Intuitions and Varieties of Ethical Naturalism', 16–7.

Similarly, Michael Huemer argues that the circularity objection utilizes an overly-generous conception of begging the question.⁴⁰ It is not the case, Huemer argues, that a writer begs the question just because she employs a premise that directly and obviously contradicts an opponent's position. Edmund Gettier's argument employs a premise that directly and obviously contradicts a certain theory of knowledge. Gettier's argument has the structure: 1) X is a case of true justified belief; 2) X is not a case of knowledge; C₁) Therefore, true justified belief does not have the same truth conditions as knowledge. Yet, Huemer argues, it would be a weak objection to claim that Gettier's counterexample begs the question against the theory that knowledge has the same truth conditions as justified true belief. We do not interpret Gettier as assuming the truth of premise 2, rather, we understand Gettier to be using his illustration to elicit an intuition that provides evidence for the truth of premise 2, evidence which, if not sufficiently opposed, licences the conclusion that X is not a case where the agent has knowledge. Similarly, Moore need not be merely assuming the truth of premise 2. Rather, Moore is relying on a linguistic intuition which, if not sufficiently opposed, licences the conclusion that D does not have the same truth conditions as goodness.^{41 / 42}

⁴⁰ Huemer, *Ethical Intuitionism*, 69–70.

⁴¹ The structure of this disagreement arises elsewhere. For example, Michael Smith argues that to cite a case of an amoralist without justification begs the question against the theory of moral judgment internalism about motivation: Smith, *The Moral Problem*, 68–70. Huemer's point may be put to Smith in this way too. Indeed, David O. Brink does reply to Smith in this way, contending that he does not think he is begging the question, rather he is offering what he takes to be a counterexample. Brink adds that he does not think he has ever simply cited the possibility of an amoralist without further argument or explanation. Brink, 'Moral Motivation', 22–3.

⁴² Other writers also maintain that OQA (Complex Property) merely invokes an intuition which provides evidence that the two terms have different truth conditions. See for example: Strandberg, 'In Defence of the Open Question Argument', 182; Shafer-Landau, *Moral Realism*, 57–8; Pigden, 'Desiring to Desire: Russell, Lewis, and G. E. Moore', 258; Heathwood, 'Moral and Epistemic Open-Question Arguments', 88. Chris Daly also discusses this point generally: Daly, *An Introduction to Philosophical Methods*, 60. One of Frank Jackson's two objections to OQA is that: if a definition of a moral term in non-moral terms survives debate and critical reflection and yet still appears to be 'open', then we are "entitled to dig in our

Whether or not Moore should be interpreted in this way, I doubt that the kind of linguistic intuitions that Ball relies on can be used as evidence against a purported definition of *G*, even evidence of relatively weak probative value. Ball acknowledges that pre-philosophic intuitions can be false, as grammatical intuitions can be, and he defends their use as evidence against two further concerns for the reliability of pre-philosophical intuitions generally. However, even if the use of pre-philosophical intuitions is standardly defensible, the more urgent problem for Ball's argument is that their use in the context of Sub-Argument for premise **(2)** lacks justification. We can see this by considering that when a rational agent, Roland, initially accepts a definition of *G* as *D*, either he is sure of the truth of the definition or he is not. If he is sure of its truth due to extensive and rigorous reflection on the matter then, along with Frankena, I doubt that he would have the linguistic intuition that something could be predicated with *D* but not *G*. Roland has done the work justifying the conclusion that the two predicates have the same truth conditions; surface intuitions offering pro tanto evidence at this point are irrelevant and unpersuasive. If someone said to Roland 'but it doesn't sound like *G* means *D*, so how can you be sure?' Roland may simply point to his reasons for believing the definition is true.

None of this is to say that Roland, or any rational agent, will no longer be open to the possibility that his conclusion is in error even after extensive, rigorous reflection. Indeed, it is a virtue to be open in this way to new ideas and challenges to one's conclusions. But being open to the possibility of an error in one's own arguments in this way depends on the kind of challenge being put to it and a mere linguistic intuition arising from the thought that

heels and insist that the idea that what fits the bill *that* well might still fail to be rightness, is nothing more than a hangover from the platonist conception that the meaning of a term like 'right' is somehow a matter of its picking out, or being mysteriously attached to, the form of the right.' So Jackson may interpret OQA according to this response, but nevertheless think that the intuition elicited is overridden. Jackson, *From Metaphysics to Ethics*, 151 (emphasis present).

one word does not sound like another should not concern Roland or someone who has more substantive arguments in favour of it.

Alternatively, Roland only tentatively accepts that G can be defined as D, in which case, there will be an explanation for his lack of confidence. Perhaps Roland only tentatively accepts the definition because he has considered all the positive arguments for it but he is yet to turn his mind to arguments against defining G as D. Alternatively, perhaps he has turned his mind to such arguments but he is yet to consider these arguments in detail. Finally, perhaps he has considered all the arguments in detail but he is yet to think about potential counterexamples. Whatever considerations Roland is yet to think through, it is his knowing that he is yet to think them through that is the source of his doubt about the truth of the definition he tentatively accepts. We know this because, as we saw above, if Roland *were* sure that the definition is true then a linguistic intuition disavouring the definition would not concern him.

Accordingly, if Roland has a linguistic intuition disavouring the truth of the definition at all, it only arises because he already has reason to doubt the truth of defining G as D. If this is right, however, then the linguistic intuition in itself could not bear any positive probative value as evidence for the falsity of the definition because the intuition itself is not doing any work, it is merely a function of the reasons Roland already has (at least at this point) to doubt the definition's truth.

It might be thought that a linguistic intuition generated in Sub-Argument for premise (2) just is one of the intuitions one comes to when thinking of potential counterexamples to the definition proposed. That is, in contrast to thinking of the linguistic intuition in Sub-Argument for premise (2) as merely a function of the reasons one already has to doubt the definition (at least at this point), it may be argued that this linguistic intuition just is a

result of substantive considerations that determine the truth of the definition, namely that the definition faces a counterexample. Yet intuitions deriving from counterexamples to a definition and linguistic intuitions deriving from Sub-Argument for premise (2) in OQA (Complex Property) differ. An intuition generated by a potential counterexample is an intuition that ‘in circumstances C performing a certain act, Φ -ing, would be G but not D’. In contrast, the linguistic intuition elicited in Sub-Argument for premise (2) consists in the thought that Question A may be answered affirmatively where Questions A & B state:

- **Question A:** Can a subject, χ , that is D, fail to be G?
- **Question B:** Can a subject, χ , that is D, fail to be D?

For the linguistic intuition derived from Question A, the details of the circumstances and the relevant act are not part of the object of the intuition. Rather, the intuition elicited is just that one thinks there *could be* a subject χ (circumstances in which a certain act is performed) that is G but not D. Accordingly, the linguistic intuition elicited in Sub-Argument for premise (2) differs to intuitions arising from concrete counterexamples. The former intuition is not responsive to a clearly defined set of circumstances. Instead it is dependent on a rational agent’s own doubts as to the possibility of a counterexample to the definition in question which, in turn, is based on the agent’s knowledge of the considerations that they are yet to turn their mind to. If the agent turns her mind to such considerations and becomes sure that the definition is true then the relevant linguistic intuition does not arise. Since the intuition derived from Sub-Argument for premise (2) is accordingly just a function of more substantive considerations about the definition, only these substantive considerations should be awarded evidential weight.

B. The Knowledge Argument

An analogous objection of circularity is put to The Knowledge Argument. Daniel C. Dennett argues that if Mary really *does* know every physical fact before she leaves the black-and-white room, then she really *may* leave the room and not learn anything.⁴³ Insofar as the image Jackson's thought experiment generates is one in which one *merely* imagines Mary knows 'lots and lots'⁴⁴ or merely all there is to know today about neurophysiology, Dennett's clarifies that 'that's just a drop in the bucket, and it's not surprising that Mary would learn something if *that* were all she knew.'⁴⁵ Mary accordingly would *not* be surprised by her experience of a blue banana upon being released, but she may be surprised at the attempt to trick her.

Dennett's objection denies step two in Sub-Argument for premise **II** and in this regard his objection is in parallel with Frankena's own concern for OQA (Complex Property)—if G really did mean 'personal autonomy', for example, then G would have the same meaning as personal autonomy, contrary to premise **(2)** in OQA (Complex Property). Dennett's conclusion is also on a par with Frankena's:

My point is not that my way of telling the rest of the story proves that Mary *doesn't* learn anything, but that the usual way of imagining the story doesn't *prove* that she *does*. It doesn't prove anything; it simply pumps the intuition that she does ("it seems just obvious") by lulling you into imagining something other than what the premises require.⁴⁶

In response to Dennett, Howard Robinson and Dale Jacquette argue that Mary's knowing all the physical scientific facts about the world does not mean that she will know of all facts because it does not provide her access to

⁴³ Dennett, "'Epiphenomenal' Qualia?".

⁴⁴ *Ibid.*, 60.

⁴⁵ *Id.* (emphasis present).

⁴⁶ Dennett, "'Epiphenomenal' Qualia?", 61 (emphasis present).

a different *type* of knowledge. Thus, Robinson argues that ‘She may know what effect a yellow object would have on her nervous system but she could not tell by looking at the object that it was having that effect.’⁴⁷ Similarly, Jacquette argues that what Mary learns is not properly tested by determining whether she correctly identifies the colour of a banana because what she learns cannot be decisively tested from an external point of view at all. The type of knowledge she acquires is only understandable from the first-person point of view for what she learns is ‘what it is like to be a color experienter, what it means to be acquainted with or to experience color qualia first-hand’.⁴⁸ Both responses to Dennett deny that the scope of the first step in Sub-Argument for premise II of The Knowledge Argument is as extensive as Dennett believes it is.

Does this response to Dennett translate into a possible reply to Frankena? On this point, I doubt it. Analogously, it may be argued that Roland’s acceptance of the definition of G as D in step (i) of Sub-Argument for premise (2) of OQA (Complex Property) is not as reliable in accounting for the whole truth of the matter as we might have thought, irrespective of his arguments in its favour. Indeed, it may be added, the linguistic intuition against the definition derived from step (ii) of Sub-Argument for premise (2), OQA (Complex Property) provides evidence of this. But whether this is so or not depends on what Roland’s arguments are and how far removed the meaning of his definiens is from the meaning of G.⁴⁹ If his arguments are sufficiently substantive and the definiens sufficiently close, his arguments will stave off any concern a mere linguistic intuition may give rise to, if any,

⁴⁷ Robinson, ‘Dennett on the Knowledge Argument’, 70–1.

⁴⁸ Jacquette, ‘The Blue Banana Trick’, 225–6.

⁴⁹ Derek Parfit, Richard Joyce (in regard to normative reasons), Allan Gibbard and Mark Schroeder, among others, make this latter point about analyses and reductions that are far removed: Parfit, ‘Reasons and Motivation’, 122; Joyce, *The Myth of Morality*, 85; Gibbard, *Thinking How to Live*, chap. 2; Schroeder, ‘Realism and Reduction’, 4–5.

as already noted. Accordingly, for most purported definitions of G, Frankena's objection to OQA (Complex Property) is sound.

IV Objection 2: Invalid Inference

A. OQA

In contrast to Frankena's objection to the truth of step (ii) of Sub-Argument for premise (2), the objection from the paradox of analysis against OQA (Complex Property) claims that either Sub-Argument for premise (2) implies a paradoxical theory of conceptual analysis or it is invalid. Sub-Argument for premise (2) is thought to imply a paradoxical theory of conceptual analysis because the linguistic test in the second step of Sub-Argument for premise (2) rules out any purported analysis of a concept if the analysans is not already known to have the same truth conditions as the analysandum. For Sub-Argument for premise (2) to be valid then, it is thought to imply that correct conceptual analyses are known in virtue of knowing the truth conditions of the concept itself. Yet successful conceptual analyses are also supposed to be informative about the term analysed (the analysandum) by describing it in certain enlightening terms (the analysans). Accordingly, it is objected that Sub-Argument for premise (2) requires successful conceptual analyses to be both: (a) uninformative; and (b) informative. And it is argued that this is a paradoxical and therefore implausible account of analyses.⁵⁰

⁵⁰ A number of writers put forward this objection including: Smith, *The Moral Problem*, 38; Jackson, *From Metaphysics to Ethics*, 151. David Lewis implicitly makes this objection by noting, in response to Moore, that analyses may be unobvious. Lewis, 'Dispositional Theories of Value', 129–30. W. D. Ross also argues that the definition of G may be true even though we are not clearly conscious of this, however, he does not put this point forward to answer a worry about a paradox of analysis. Ross, *The Right and the Good*, 93.

In particular, it is claimed that this view is problematic for cases where one is unaware of a tautology that others (such as Mary) are aware of. For example, Moore's test in premise (1) may posit the following two questions:

- "Can a subject, χ , that is *mauvais* fail to be bad?
- "Can a subject, χ , that is *mauvais* fail to be *mauvais*?

'Bad' translates as '*mauvais*' in French, and so Mary would know that both questions have trivial answers. Yet for one unaware of this, the linguistic test in Sub-Argument for premise (2) demonstrates that the meaning of each predicate is different. A similar problem arises for analyses of concepts (possibly including G) that we are presently unaware are correct.⁵¹

Instead, it may be insisted that conceptual analysis does not require that the analysans already be consciously known:

Why are analyses unobvious and informative? Because even though someone who has a mastery of some concept C must *have* certain inferential and judgemental dispositions, it may not be transparent to her what these inferential and judgemental dispositions are, and so, *a fortiori*, it need not be transparent to her what the best summary or systematization of the platitudes that describe these dispositions is. Whereas mastery of a concept requires knowledge-how, knowledge of an analysis of a mastered concept requires us to have knowledge-that about our knowledge-how.⁵²

For example, a vociferous, amateur Los Angeles Lakers fan has an intuitive understanding of the game. She knows when a player has travelled or

⁵¹ This is essentially Alan H. Goldman's response to a version of OQA levelled at reasons, although Goldman does not consider the argument in any detail. Goldman, 'Desire Based Reasons and Reasons for Desires', 481. Furthermore, the objection here can be taken further as Thomas Baldwin observes, given that, this objection is not only critical of Moore's method of analysis, but it also refutes Moore's conclusion that G is indefinable. Even if G is indefinable, by Moore's own lights this is either trivial or false, for if G were definable then we would trivially observe the analysans within G itself and since we cannot, its indefinability is trivial. Baldwin, *G.E. Moore*, chap. 3, sec. 5.

⁵² Smith, *The Moral Problem*, 38 (emphasis present).

charged, yet she cannot state the specific rules referees follow when deciding such matters. The amateur fan has knowledge-how (she knows how to correctly apply her concepts in most cases), but presently lacks 'knowledge-that' about her knowledge how; she lacks the ability to articulate the precise theory underlying her 'knowledge-how'.⁵³ This understanding of conceptual analysis therefore avoids the purported paradox by rejecting Moore's requirement that analyses be transparent and hence, uninformative. Instead, it allows that knowledge-that about our knowledge-how may be opaque.

Ultimately then, even if a predicate's definition is not known in virtue of knowing the meaning of the predicate itself, or known in this way at first blush, it may nevertheless be true (and indeed become more well known) as when one learns the meaning of *mauvais* or the rules of travelling.⁵⁴ But if we accept that correct conceptual analyses can be opaque in this way then it does not follow from the mere linguistic intuition elicited in Sub-Argument for premise (2) that the analysis is false. This means that Sub-Argument for premise (2) is invalid and, without more, that premise (2) need not be true.

This objection has been met with the straightforward reply that Sub-Argument for premise (2) does not imply that correct analyses be both: a)

⁵³ The distinction is initially introduced by Gilbert Ryle: Ryle, *The Concept of Mind*, 2.

⁵⁴ So defined, the puzzle of explaining how successful conceptual analyses can be both informative and knowable a priori resembles both Meno's paradox of inquiry and, according to Hans-Johann Glock and John Skorupski's interpretation, a problem for Kant. Meno's paradox of inquiry is that: for one to know something that one does not know, one must not know it (for there to be a need to search for it) yet one must also know it already (to be able to recognise what to search for): Plato, 'Meno', p.200, 80d–e. In comparing the two, Nicholas White observes that while in both puzzles a certain effort is unnecessary if it is possible, on the standard interpretation of the paradox of analysis the puzzle arises in regard to conditions for an analysis. In contrast, Meno's paradox of inquiry arises in regard to recognizing a successful analysis: White, 'Inquiry', 302–4. According to Glock and Skorupski, Kant's project required him to explain how knowledge could be synthetic a priori: how it could both tell us something about reality ('synthetic'), and yet also be knowable independently of experience (a priori): Glock, *What Is Analytic Philosophy?*, chap. 1; Skorupski, *The Domain of Reasons*, 166.

uninformative; and **b**) informative. Indeed this reply follows from Ball's response to the circularity objection. As the linguistic test in Sub-Argument for premise (2) merely provides evidence that an analysis of the concept of G is unsuccessful, it is claimed that it need not be inferred that OQA (Complex Property) implies that successful analyses are known to be true in virtue of knowing the meaning of the predicate itself. A native English speaker may form the intuition that D is not G, without presupposing that G does not mean D.

Since this reply to the objection from the paradox of analysis uses the same response employed to avoid the circularity objection, it is also vulnerable to the objection above. The linguistic intuition elicited in Sub-Argument for premise (2) is not available for use as evidence against the truth of a proposed definition of G. Accordingly, the dilemma presented in this objection to OQA (Complex Property) cannot be avoided in this way.

B. The Knowledge Argument

Analogously, Lawrence Nemirow and David Lewis, among others, deny that Mary acquires non-physical knowledge after leaving the black-and-white room. Instead, they assert, Mary acquires an *ability* to remember, recognize and imagine sensing red after her first experience of doing so:⁵⁵

These abilities to remember and imagine and recognize [tasting Vegemite] are abilities you cannot gain (unless by super-neurosurgery, or by magic) except by tasting Vegemite and learning what it's like. [...] The ability hypothesis says that knowing what an experience is like just *is* the possession of these abilities to remember, imagine, and recognize. It isn't the possession of any kind of information, ordinary or peculiar. It isn't knowing that certain possibilities aren't actualized. It isn't know-that. It's knowing-

⁵⁵ Lewis, 'What Experience Teaches'; Nemirow, '*Mortal Questions*, by Thomas Nagel', 475–6.

how. [...] Lessons impart information; ability is something else. Knowledge-that does not automatically provide know-how.⁵⁶

Lewis' argument accordingly denies that Sub-Argument for premise II is valid. He agrees that Mary has complete physical knowledge while she is in the black-and-white room and he agrees that Mary learns something about the color experiences of others upon leaving the room. However, since what Mary learns is an ability (and hence not a physical fact(s)), Lewis denies that the truth of premise II follows from these two propositions. Mary can have complete knowledge and still learn something: an ability.

Accordingly, both this objection to The Knowledge Argument and the paradox of analysis objection to OQA (Complex Property) maintain that the relevant intuition elicited to support the denial of the consequent fails. In each case a reliance on the intuition assumes that one type of knowledge (know-that cf. also know-how)⁵⁷ or one mode of believing (intellectual cf. also practical)⁵⁸ exists. In response to OQA (Complex Property) it is argued that we have know-how about our concepts and we acquire know-that about our know-how. In response to The Knowledge Argument it is argued that Mary has know-that about other people's experience of colour and what she acquires is knowledge of how to remember, imagine and recognize that experience.

This objection has faced significant criticism, yet in this regard also, the concern with a linguistic intuition in OQA (Complex Property) compared to an epistemic intuition in The Knowledge means that such criticism is largely inapplicable to OQA (Complex Property). Thus, William Lycan denies the tenability of the know-how/ know-that distinction because he considers that

⁵⁶ Lewis, 'What Experience Teaches', 99–100 (emphasis present).

⁵⁷ Smith, *The Moral Problem*; Lewis, 'What Experience Teaches'.

⁵⁸ Pettit, 'Motion Blindness and the Knowledge Argument'.

know-how *is* know-that.⁵⁹ Other writers also comment on this purported distinction yet, for the purposes of OQA (Complex Property), we can see that the outcome of this issue is irrelevant because even if Lycan is right, a proponent of the paradox of analysis objection may simply declare that a competent speaker does not have knowledge-that in relation to the semantic equivalence of two predicates. It is presupposed in Mary's case that she knows all physical facts. No similar epistemic presupposition is made in OQA (Complex Property) and so the tenability of a distinction between two types of knowledge has no bearing on the objection.

It is also argued in response to Lewis that Mary must acquire more than an ability. For example, Martine Nida-Rümelin argues that upon seeing a certain colour for the first time, but not yet knowing its name, (let us say she sees something blue), Mary's *ability* to remember, imagine, and recognize that colour may explain what she learns when she sees it for the first time. However, Nida-Rümelin maintains, this ability does not explain Mary's learning that *this* colour (and not, for example, the orange colour she also sees) is what we call 'blue'. Thus, Mary must also acquire knowledge of a truth or, 'phenomenal knowledge'.⁶⁰ For OQA (Complex Property), analogously one might argue that from the linguistic test we acquire more than mere evidence of semantic inequivalence, we acquire the truth of such inequivalence. Yet, as discussed, epistemic inequivalence need not entail semantic inequivalence.

Finally, Michael Tye argues that the explanation of the invalid inference cannot be that Mary acquires a certain ability in the sense Lewis claims because when Mary sees a red rose for the first time, she sees a particular shade of red, 'red17' for example, and yet it is not true that she acquires the

⁵⁹ Lycan, *Consciousness and Experience*, 92–4.

⁶⁰ Nida-Rümelin, 'What Mary Couldn't Know: Belief About Phenomenal States', 258.

ability to remember and imagine red.⁶¹ At best Lewis can claim that Mary acquires this ability for as long as she is observing the rose. Yet, Tye argues, this cannot be right either because Mary may observe the rose for the first time while distracted or while thinking of something else altogether. In this case Mary does not acquire knowledge of the experience of the redness¹⁷ of the rose even while observing it, yet she does have the ability to ‘mentally point to the phenomenal character of her experience with an indexical concept via introspection.’⁶²

In this case as well, an analogous response to the objection from the paradox of analysis is unavailable. Analogously, one could claim that the explanation of the invalid inference in OQA (Complex Property) cannot be that the competent speaker lacks knowledge-that about his knowledge of how to use certain predicates because he may not remember the knowledge-that he acquires or he may not be thinking about it at the time the justification for semantic equivalence is being explained to him. Yet because OQA (Complex Property) is concerned with semantic equivalence and not the competent speaker’s knowledge we can still consistently explain the speaker’s invalid inference as arising from the speaker’s lack of knowledge, or recall, or his inattentiveness while listening to the justification for semantic equivalence.

V Objection 3: Property Identification

A. OQA

The objection from property identification maintains that while the linguistic test in Sub-argument for premise (2) in OQA (Complex Property) may be

⁶¹ Tye, ‘Knowing What It Is Like: The Ability Hypothesis and the Knowledge Argument’.

⁶² *Ibid.*, 153–4.

relevant to defining G , it is irrelevant, or largely irrelevant, to the question of whether G_R is reducible to another property. It therefore disputes premise (4) in OQA (Referent).⁶³

The objection has multiple forms. One version derives from Gottlob Frege's distinction between a conventional sign, the relatively objective sense that a sign expresses, and the object in the world, if any, which the sense designates—its referent.⁶⁴ According to Frege, identity relationships of the form " $a = b$ " (cf. " $a = a$ ") may be true and informative because, even though the sense of the sign " a " may differ to the sense of the sign " b ", each sign may designate the same referent. In this context, Frege describes the sense of a sign as the bearer of 'the mode of presentation'⁶⁵ of the referent and as being 'grasped by everybody who is sufficiently familiar with the language'.⁶⁶ While in some qualified form the sense does convey a meaning of its sign, this meaning is not necessarily the same as the sign's referent. The referent or denotation is instead the object that the sense designates. Thus, the two singular terms ' 6^3 ' and ' 216 ' each possess a different sense despite designating the same number. Indeed, Hilary Putnam notes that this is generally the case with an ostensively learned term for a property such as 'has high temperature': 'it takes empirical and theoretical research, not linguistic analysis, to find out what temperature is (and, some philosopher

⁶³ This objection at least dates back to 1970: Durrant, 'The Identity of Properties and the Definition of "Good"'. Other proponents of this objection include: Putnam, *Reason, Truth and History*, 208; Brink, 'Realism, Naturalism and Moral Semantics'; Sturgeon, 'Moore on Ethical Naturalism', 534; Schroeder, 'Realism and Reduction', 3–4, 10 n26; Schroeder, *Slaves of the Passions*, 65, 72–3.

⁶⁴ Frege, 'On Sense and Reference', 57. Frege restricts the application of this account of identity to 'any designation representing a proper name, which thus has as its reference a definite object (this word taken in the widest range)' and expressly excludes concepts and relations from this class to which he refers the reader to another paper. As Moore considers G_R to be simple in the sense of having no parts it is unlikely Moore thought of G or G_R as a 'concept' in the sense Frege was wary of. However, Frege's qualification may be relevant for alternative interpretations of G and G_R .

⁶⁵ Id.

⁶⁶ Ibid., 57–58.

might suggest, what *goodness* is)'.⁶⁷ If from empirical research it is discovered that, for example, the property of 'temperature' is identical to the property of 'mean kinetic molecular energy', we have what Putnam describes as a synthetic identity of properties.⁶⁸ Importantly then, this identity relation may obtain even where the sense of each term varies. Putnam accordingly argues that Moore's implicit denial of this identity relation is 'devastating',⁶⁹ perhaps especially so for synthetic identities that obtain necessarily. Similarly then, it may be argued that even though the sense of G is different to the sense of its purported definiens, the two senses co-refer to a certain (natural, non-natural or supernatural) referent, contrary to premise (4), OQA (Referent).⁷⁰

Michael Smith recently denies this possibility.⁷¹ Smith argues that proponents of the property identification objection claim that determining what the properties of G and rightness are, is an a posteriori truth because we can use the predicates 'G' and 'rightness' to explain certain empirical phenomena. Based on such observations, we can fix the reference of G via a description of the effects of actions which we invariably invoke the predicate G to explain. We may, for example, observe that all the conduct in which we infer that G_R is present is causally responsible for a tendency towards social stability. In light of this, we can fix the reference for G as 'the property, whatever it is, that is causally responsible for a tendency towards social stability.' With this reference in hand, we can examine acts with this effect to determine what property or properties explain this tendency. Suppose we

⁶⁷ Putnam, *Reason, Truth and History*, 207.

⁶⁸ *Ibid.*, 85.

⁶⁹ *Ibid.*, 208.

⁷⁰ Millian or Russellian theories of semantics may be introduced instead of Fregean theories to answer OQA (Complex Property) in a relevantly similar way.

⁷¹ Smith, 'Moral Realism', 197–8. Variations of this argument are given by Terence Horgan & Mark Timmons, and more recently, Susana Nuccetelli & Gary Seay: Horgan and Timmons, 'New Wave Moral Realism Meets Moral Twin Earth'; Horgan and Timmons, 'Troubles for New Wave Moral Semantics'; Nuccetelli and Seay, 'What's Right with the Open Question Argument', 278–80.

discover that the relevant property is the property 'promoting the well-being of sentient creatures', then we can conclude that G_R is identical to this property and this determination will be a posteriori, not a priori, and hence, not subject to the objection that the sense of G differs from the sense of 'promoting the well-being of sentient creatures'.

Smith agrees that the natural moral realist's conclusion here is determined a posteriori and to this extent it avoids OQA (Complex Property). However, he contends that in fixing the reference of G via the description 'the property, whatever it is, that is causally responsible for a tendency towards social stability', a proponent of this objection is defining G a priori.⁷² Accordingly, Smith contends, OQA (Complex Property) can be put to a natural moral realist at the point at which she provides a reference-fixing description of G and hence, before she can conduct an a posteriori investigation into what G_R consists in.

Smith is largely right, but as it stands his objection is open to reply. As Smith maintains, without firstly delimiting the circumstances in which G_R is known to exist, one may identify any property with G_R on an a posteriori basis. However, it may be argued that the reference of G can be fixed in a rather trivial way. It may, for example, be fixed by the description 'that property, whatever it is, that is present in acts that are G '. Evidently, it will be controversial what those acts are. But to avoid dissension, investigation could be restricted to acts that are considered paradigmatically G . Isolating cases that are paradigmatically G would be different to fixing the reference of G itself. The reference is fixed by the rather trivial description already noted. The method of investigating instances of G_R in the world need not be restricted in the same way. Indeed it cannot be. For, if the reference 'air

⁷² It is true, I may add, that this definition of G is based on the realist's observation of conduct, but a conclusion will nevertheless still be determined a priori if it is considered to be true independently of experience. Arguably this is the case here.

breathable by humans during a certain period of time, d' is fixed as 'the airy stuff in the world breathable by humans during d' ', investigating the properties that play this role does not require examining every sample of airy stuff in the world during d . Rather, we would turn to examine samples from, for example, the air we are currently breathing and we will eschew examining samples from, for example, sandstorms, billowing smoke stacks and volcanic air pockets. No doubt there will be variation in the samples collected, but we would only be interested in those components common to each. With sufficient data in hand, consistently with the scientific method, we can infer that our results generalize to other instances of airy stuff in the world breathable by humans during d . Analogously, we can confine our investigation of G_R in the world to acts paradigmatically bearing this property.

However, when investigating *moral* and normative properties in this way, the rather trivial reference-fixing description must be more discerning. Even if we investigate only acts that are paradigmatically G (and to be careful, acts that are not: paradigmatically morally evil, paradigmatically non-moral, paradigmatically morally good for some people, paradigmatically morally evil for some people, paradigmatically non-moral for some people), given the trivial reference-fixing description of G that is our starting point, we would have no way of determining which one, or more, of a number of common properties is identical to G_R . That is, G_R would remain indeterminate in virtue of an ambiguous reference-fixing description. Perhaps most starkly, such ambiguity arises because, for moral predicates, an investigator will have to front up to the question of the extent, if any, to which internal states in moral agents matter. Thus, for example, one philosopher might regard Payton's helping an injured pedestrian as morally good, irrespective of why she provided such assistance and how she felt in doing so. Another philosopher, alternatively, may regard Payton's assistance as morally good, not only because of the aid she provided but because she

believed it would help, and/ or because she felt appropriately sympathetic for the pedestrian's plight. In such circumstances, and for other paradigmatically morally good acts, an investigator has no way of identifying which property or properties are identifiable with G_R without going beyond the initial trivial reference-fixing description 'that property present in acts that are G '. Yet, in leaving aside the trivial reference-fixing description to provide a more discerning reference-fixing description, the description must face OQA (Complex Property). Furthermore, there is no reason to believe that a question about the relevance of moral agents' mental and affective states is the only one an investigator must answer. The upshot is that there will be a plurality of properties common to acts described as paradigmatically G where G_R is fixed by a trivial reference-fixing description and investigators will have to explain why one such property rather than another, or a combination, is identical to G_R .

It might be replied that none of this means that G_R is not identical to some other property, of any kind (normative, non-normative, natural etc.) that we presently lack the vocabulary to describe.⁷³ I think this is right. G_R might be reducible to property ____ or the properties: 'promoting personal liberty' and ____, even if we have no idea what natural property this is or what kind of natural property it is. However, if a property in G_R 's reductive base is without a name and we are without an idea of what *kind* of natural property it is, then investigators would be just as unable to identify it a posteriori as they would be unable to, were G given a trivial reference-fixing description. Indeed, investigators could only proceed via a trivial reference-fixing description. It might be explained that property ____—whatever it is—is a

⁷³ Mark Schroeder makes this observation: 'Suppose that being good was just contributing to overall happiness, but that no one had ever coined the word "happiness", and so there was no descriptive predicate, "contributing to overall happiness" to express the property of *goodness*. Would it follow that the property of *goodness* was irreducible?' Schroeder, 'Realism and Reduction', 10 n26 (emphasis present).

kind of affective state, but specifying the kind of property ____ is moves beyond providing a trivial referencing-fixing description. The only alternative is for a proponent of this view to insist that our inability to determine which natural property G_R is reducible to does not mean that G_R is irreducible. Again this is true, but it leaves neither party with anything else to say, except perhaps for our wondering what this proponent's motivation is for believing that such a reduction obtains, and hence, why her adherence to the possible reducibility of G_R in this way is not *ad hoc*. She may have an answer, but without providing more than a trivial definition of G , I am not sure what this could be.

Instead then of relying on a trivial reference-fixing description or no description, proponents of the property identification objection are left to employ a non-trivial reference-fixing description of G , perhaps devised according to process of broad reflective equilibrium as normative ethicists do and as one proponent of the property identification objection counsels.⁷⁴ Of course, upon fixing the reference of G in this way, the relevant description will be vulnerable to OQA (Complex Property), as Smith suggests. The disanalogy with investigating samples of air and non-moral properties is that it is far less plausible that one of our own internal states (*believing* that something is air breathable by humans during period of time d , for example) is part of what makes a gas 'air breathable by humans during d ', among other things. For moral properties, investigators must first face OQA

⁷⁴ David O. Brink argues that a causal theory of reference should not be understood as referring just to properties that regulate the terms people actually use, but also as referring to properties regulating the terms they would use upon proper reflection of imagined situations. And, what is required, Brink explains, for proper reflection is broad 'dialectical equilibrium' where he explains this is similar to broad reflective equilibrium. It is in virtue of such a process that Brink claims that we can identify what properties causally regulate our use of moral terms. Ultimately, Brink eschews a causal theory of reference in favour of a shared intention account of reference, yet again, he thinks that moral reasoning is critical to fixing the referent and indeed he considers it an objection to a direct reference theory that it renders moral reasoning obsolete. Brink, 'Realism, Naturalism and Moral Semantics', 168–9.

(Complex Property), if only by contending that it is unsound for some other reason.

B. The Knowledge Argument

A similar objection plagues The Knowledge Argument. It maintains that while Mary *does* learn a fact or truth after leaving the room, the knowledge she acquires is simply knowledge of certain physical properties she already knew about, described in experiential terms. Thus, Brian Loar, for example, argues that the phenomenal concept Mary acquires after seeing red for the first time is separate from her physical-functional concepts. However, Loar explains, the two concept words can co-refer to one physical property that Mary did know about.⁷⁵

One standard objection to this response is that it just relocates the problem because one fact Mary does learn is: that her new phenomenal concept and her physical-functional concept are ways of being given a particular physical property.⁷⁶ Loar attempts to answer this objection in his account⁷⁷ but, for our purposes, the point is that this response translates into a second plausible objection to the sense-reference objection to OQA (Complex Property).

We have followed Smith in accepting that the proponent of the property identification objection has a problem in determining by a posteriori investigation that the referents of G and another predicate are identical because she has to fix the description of G first. Analogously, to the standard objection to Loar's response, just noted, we can see that even if the problem

⁷⁵ Loar, 'Phenomenal States (Revised)', 222–5.

⁷⁶ For example: Chalmers, 'Phenomenal Concepts and the Knowledge Argument', 287–90.

⁷⁷ Loar, 'Phenomenal States (Revised)'. David Chalmers also responds to Loar's answer to this objection: Chalmers, 'Phenomenal Concepts and the Knowledge Argument', 290–3.

Smith raises is avoidable, a variant of OQA (Complex Property) re-emerges. The argument is that appeal to the thought that 'G and another predicate have different senses but nevertheless co-refer' only relocates the problem, because the linguistic test in Sub-Argument for premise (2) can be given in regard to the property that each predicate purportedly refers to. That is, it can be asked:

- **Question A:** Can a property D, referred to by predicate D, fail to be referred to by predicate G?
- **Question B:** Can a property D, referred to by predicate D, fail to be referred to by predicate D?⁷⁸

VI Objection 4: Invalidity

Section IV above argued that Sub-Argument for premise (2) in OQA (Complex Property) and Sub-Argument for premise II in The Knowledge Argument were invalid. In contrast, Gilbert Harman argues that OQA (Complex Property) is invalid in its entirety⁷⁹ and Stuart Brock and Edwin Mares formulate an analogous objection in regard to color.⁸⁰ However, ultimately Harman's objection is a version of either the objection from the paradox of analysis or the property identification objection. It can collapse into either because Harman's statement of Moore's OQA and criticism of it does not distinguish between semantic and ontological levels.⁸¹

⁷⁸ On this point also see: Skorupski, *The Domain of Reasons*, 447–8.

⁷⁹ Harman, *The Nature of Morality*, chap. 2.

⁸⁰ Brock and Mares, *Realism and Anti-Realism*, 104.

⁸¹ This would be because Harman puts forward this criticism in an introductory text and so the omission of this detail may have been wise.

VII Conclusion

Further objections are also put to OQA (Complex Property),⁸² however it is sufficient to have identified problems of vicious circularity and invalid inference. Despite this, realists and anti-realists often mention that there is more to the matter than this.⁸³ Some scholars have accordingly sought to trace the source of the open ‘feel’ of G with a view to successfully accounting for this feature in a naturalistic definition.⁸⁴ Conversely, others have sought to identify the source of the openness of G with a view to reconstructing Moore’s argument and proving that G is *indefinable*. In this mode, A. J. Ayer argues that moral concepts are unanalysable because they are merely ‘pseudo-concepts’, adding nothing more to the factual content of non-moral propositions than an expression of moral approval or disapproval.⁸⁵ R. M. Hare instead emphasises the special function of commending and condemning of value terms, a function that entails a reason for action once assented to.⁸⁶ J. L. Mackie, in contrast, explains that it is because the ‘practicality’ of moral judgments is wholly relative to an agent’s desires or possible satisfactions that naturalistic definitions of moral terms are unable

⁸² See for example: Peter Geach on the attributive use of ‘good’, E. Clark and M. Clark on disambiguating ‘good’, Peter Railton on reformative definitions and Nicholas Sturgeon, firstly on Moore’s generalization and secondly, on Moore’s question-begging assumption that G must be defined in non-ethical terms to be a natural term itself and refer to a natural property itself: Geach, ‘Good and Evil’; Clark and Clark, ‘What Is Goodness?’; Railton, ‘Naturalism and Prescriptivity’, 157–9; Sturgeon, ‘Moore on Ethical Naturalism’, 535, 535–40.

⁸³ Darwall, Gibbard, and Railton, ‘Toward *Fin de siècle* Ethics: Some Trends’, 116; Blackburn, *Ruling Passions*, 86; Gibbard, *Thinking How to Live*, 22; Sayre-McCord, ‘Moral Realism [SEP]’, sec. 3.

⁸⁴ For example: Rosati, ‘Naturalism, Normativity, and the Open Question Argument’.

⁸⁵ Ayer ‘Critique of Ethics and Theology’ in *Foundations of Ethics: An Anthology*, 42.

⁸⁶ Hare, *The Language of Morals*, 91, 124–6. In regard to the relation between: **a**) evaluative judgments and **b**) prescriptions to act as well as between: **a**) assent to a prescription and **b**) having a reason to act see: *Ibid.*, 164–72; Hare, *Freedom and Reason*, 51.

to capture the categorical quality of moral judgments.⁸⁷ Several other writers similarly consider that the source of the openness of G, and all ethical terms, lies in their practical function or 'normativity', compared to the purely descriptive function of descriptive terms.⁸⁸ More recently, Connie S. Rosati and Stephen Darwall argue that in addition to this, the open feel of moral terms must also be at least partly explained by an agent's capacity for autonomous evaluation and action, which is constitutive of their agency.⁸⁹ In Chapters 3-6 we turn to an argument maintaining that G is indefinable in light of its reason-giving function.

⁸⁷ Mackie, *Ethics: Inventing Right and Wrong*, 33.

⁸⁸ Thomas Baldwin, Christine Korsgaard and Simon Blackburn cite G's characteristics of: guiding our thoughts and judgments, aiding us in identifying what is important and what to do, as explaining OQA: Baldwin, 'Editor's Introduction', xix; Baldwin, *G.E. Moore*, 89; Baldwin, 'The Indefinability of Good', 321; Baldwin, 'A Hundred Years of *Principia Ethica*', 10; Korsgaard, *The Sources of Normativity*, 43; Blackburn, *Ruling Passions*, 70. Donald Regan, Stephen Darwall, Paul Bloomfield and William J. Fitzpatrick all specifically cite the 'normativity' of G as explaining the force of OQA: Regan, 'How to Be a Moorean', sec. II; Darwall, 'Internalism and Agency', 160; Darwall, 'Moore, Normativity, and Intrinsic Value', secs. I-II; Bloomfield, 'Opening Questions, Following Rules'; FitzPatrick, 'Robust Ethical Realism, Non-naturalism, and Normativity', 176. Moore himself may have toyed with the idea that, at least in part, 'good' is indefinable because it is not merely describing a feature, but prescribing that a certain thing is really what we ought to do. See Charles Pigden's discussion of Moore's 'Barren Tautology Argument' in this regard: Pigden, 'Desiring to Desire: Russell, Lewis, and G. E Moore'. David Wiggins, Thomas Scanlon, and Terence Horgan & Mark Timmons also mention G's 'normativity' as explaining OQA, yet these authors flesh this out in terms of G being reason-providing: Wiggins, 'Cognitivism, Naturalism, and Normativity: A Reply to Peter Railton', 304, 305-7; Wiggins, 'A Neglected Position?'; Scanlon, *What We Owe to Each Other*, 57-8, 96-7; Horgan and Timmons, 'Moorean Moral Phenomenology', 212. In contrast, Thomas Nagel suspects that it is an unrecognized assumption of G's relation to motivation that explains Moore's supposed refutation of naturalism: Nagel, *The Possibility of Altruism*, 8.

⁸⁹ Rosati, 'Agency and the Open Question Argument'; Rosati, 'Naturalism, Normativity, and the Open Question Argument'; Darwall, 'Moore, Normativity, and Intrinsic Value'. For criticism, see: Enoch, 'Agency, Shmagency'; FitzPatrick, 'Robust Ethical Realism, Non-naturalism, and Normativity', 177, 177 n45.

3

Motivation In Intentional Conduct

While various attempts to reconstruct G. E. Moore's OQA prove to be untenable, we have seen that a number of writers consider that G is nevertheless indefinable in natural terms in light of some normative characteristic it possesses. One argument in particular that is founded on this view and is considered by some to be successful maintains that G is indefinable in light of its reason-giving function.¹ The argument is expressed in the following argument-text schema:

P₁ A moral fact entails x

P₂ All x 's are \mathcal{R}

P₃ (P₁ & P₂) entails that moral realism is false

C Moral realism is false

where ' x ' denotes a kind of entity.

¹ Richard Joyce, for example, bases his defence of moral error theory on this argument. Joyce, *The Myth of Morality*, 42. However, *pace* Joyce, I think there are three problems with his argument. Firstly, he defends Bernard Williams' version of Reason/ Motivation Internalism which is flawed. Secondly, his defence of HTM is incomplete and thirdly, his argument is invalid in virtue of assuming that a realist or objective morality must be committed to propositions that our pre-reflective concept of morality involves. Notably, Joyce recently addresses part of this third worry: Joyce, 'The Accidental Error Theorist'.

In The Relative Reason Argument, as we shall see, 'x' in this schema denotes a type of justifying reason and the predicate letter ' \mathcal{R} ' denotes 'is agent-relative'.² Explicitly, the argument is:

The Relative Reasons Argument

- P₁** A moral fact or a sincere moral judgment that an agent *A* should perform a certain act, Φ , in circumstances *C* at time *t* necessarily entails a sufficient or conclusive practical justifying reason for *A* to Φ in *C* at *t*. (*Morality/ Reason Internalism*)
- P₂** All sufficient and conclusive practical justifying reasons are agent-relative. (*Relative Reasons*)
- P₃** According to Moral Impartiality, in propositions 1-3 of Moral Realism, either:
- a)** facts exist independently of moral agents; or
- b)** where **a)** does not apply to a given entity, any relation between the entity and moral agents is agent-neutral.

C₁ Moral Impartiality is false (and hence, Moral Realism is false).

The argument is valid: if moral facts necessarily arm moral agents with agent-relative reasons to act, then it is not the case that the truth of moral propositions is independent of moral agents and nor is it the case that there is an agent-neutral relation between the truth of moral propositions and moral agents.³ The implication is that Moral Impartiality and therefore Moral

² The well-known moral sentimentalist argument from David Hume also fits this schema: Hume, *A Treatise of Human Nature*, bk. 3.1.1[1]–[6], pp.507–9. In this regard also see: Hutcheson, 'A System of Moral Philosophy, Volume I: Concerning the Constitution of Human Nature and the Supreme Good', bk. 1.1.III[2]; bk. 1.2.V[1]–[3].

³ Although disjunct **a)** in Moral Impartiality does not refer to *true* moral propositions but just moral propositions, proposition 3 (Propositional Truth) in the definition of Moral Realism establishes that at least one moral proposition is true. Accordingly, by

Realism is false. Accordingly, if the argument is unsound, either **P₁** or **P₂** must be false.

This discussion of the validity of The Relative Reasons Argument presupposes that the meaning of ‘sufficient’ or ‘conclusive practical justifying reason’ in **P₁** and **P₂** is the same and this is easily overlooked. **P₁** may readily be considered to be true under a rather neutral understanding of ‘sufficient’ and ‘conclusive practical justifying reason’ that differs to the meaning of these terms when they are later partly defined and discussed in **P₂**. To avoid potential ambiguity, I therefore consider **P₂** first.

A separate argument with the following form is sometimes put forward for this premise:

The Reason/ Motivation Argument

- | | |
|----------------------|--|
| P₄ | All sufficient and conclusive practical justifying reasons for a rational human agent, <i>A</i> , to Φ in <i>C</i> at <i>t</i> are necessarily related to <i>A</i> ’s capacity to be motivated to intentionally Φ (or omit to Φ) in <i>C</i> at <i>t</i> . (<i>Reason/ Motivation Internalism</i>) |
| P₅ | A rational human agent’s motivation to intentionally Φ (or omit to Φ) in <i>C</i> at <i>t</i> for a sufficient or conclusive subjective justifying reason is always and only correctly explained by an agent-relative explanatory reason. (<i>The Theory of Relative Motivation or ‘TRM’</i>) |
| <hr/> | |
| P₂ | All sufficient and conclusive practical justifying reasons are agent-relative. (<i>Relative Reasons</i>) |

The Reason/ Motivation Argument is also valid, but **P₄** and **P₅** are contentious. This chapter focuses on **P₅**. This premise is usually stated as a version of The Humean Theory of Motivation (hereafter ‘HTM’). I define HTM in Section I and, in Section II, explain why the theory is inadequate as a

implication disjuncts **a)** and **b)** in Moral Impartiality refer to true moral propositions as well.

substitute for **P₅** and why **P₅** itself should be slightly qualified. The result is a theory I call Relative & Contingently Neutral Explanations ('R&CN'). In turning to assess R&CN, in Section III I leave aside arguments based on empirical work to instead focus on identifying what can be said about human motivation in the most general terms possible, and hence, for all possible human motivation in the actual world (cf. part of it) and in close possible worlds. I also maintain that a certain set of recent arguments that might be used to assess R&CN rely on phenomenology, and are therefore unsound. Nevertheless, Section IV develops an a priori objection to R&CN.

I The Humean Theory of Motivation

HTM is stated in slightly different ways. I take it to consist in the conjunction of two principles:

The Contents Principle: an explanation of a human agent's intentional moral or non-moral conduct necessarily consists in at least two modally distinct mental states (a 'belief' and a 'desire'), which the agent has 'put together' merely in the sense that she does not presently possess the two by mere coincidence.⁴

The Original Attitude Principle: the distinct mental state of a desire in The Contents Principle is an 'original' mental state in the sense that it is not brought into existence by a belief state or a process of reasoning.

A second, less commonly stated version of HTM maintains that HTM consists solely in The Contents Principle⁵ while other writers leave the

⁴ The term *intentional* here denotes the ordinary mental state in which an agent is primarily focused on performing the relevant act and/ or the result of it. It does not denote the mental state ascribed to agents in Harman and Knobe-like cases wherein the *consequences* of the agent's act determine whether or not the agent is judged as having acted intentionally. Harman, 'Practical Reasoning'; Knobe, 'Intentional Action in Folk Psychology'. G. F. Schueler and Nikolaj Nottelman also maintain this distinction: Schueler, *Desire*, 35–6; Nottelmann, 'Belief-Desire Explanation', 913.

⁵ Smith, 'The Humean Theory of Motivation', 36; Smith, *The Moral Problem*, 92, 179; McNaughton, *Moral Vision*, 47, 106; Schueler, *Reasons and Purposes*, 22; Miller, *An*

matter unclear.^{6/ 7} However, The Original Attitude Principle denies the possibility that human motivation can be, or, is sometimes initiated by a belief(s) alone which subsequently generates the purportedly required desire, as Thomas Nagel, among others, identifies.⁸ As a result, this version more accurately depicts the theory of motivation that writers are concerned with and, indeed, most writers expressly define HTM in this way.⁹ It also more closely resembles Hume's own view.¹⁰

Introduction to Contemporary Metaethics, 7, 270–9; Skorupski, *The Domain of Reasons*, 238, 240–1; Nottelmann, 'Belief-Desire Explanation', 913.

⁶ Copp and Sobel, 'Desires, Motives, and Reasons: Scanlon's Rationalistic Moral Psychology', 245–6; Railton, 'Humean Theory of Practical Rationality', 266; Broome, 'Reasons and Motivation II', 140.

⁷ Jonathan Dancy maintains that in addition to The Contents Principle, HTM asserts that desire states are dominant in some way. Although he acknowledges that Humeans do not presently agree upon the way in which they are dominant, he explains that their dominance need not be explained by further asserting that desires are original; desires may be dominant in light of their direction of fit: Dancy, *Practical Reality*, 11–12.

⁸ Nagel, *The Possibility of Altruism*, 28–30; Nagel, *The View From Nowhere*, 151; Nagel, *The Last Word*, 102–3. Other academics have also followed this line of response: Wallace, 'How to Argue about Practical Reason', 361–371, 375ff; Smith, *The Moral Problem*, 160–1; Smith, 'In Defence of The Moral Problem: A Reply to Brink, Copp, and Sayre-McCord'; Korsgaard, 'Skepticism about Practical Reason'. James Lenman and Neil Sinhababu also interpret John McDowell in this way, however, since McDowell considers the relevant psychological state as a 'belief-entailing-an-EA' there is a stronger connection between the belief and the EA than this. In virtue of this, McDowell might more accurately be interpreted as believing instead in the existence of besires. McDowell, 'Are Moral Requirements Hypothetical Imperatives?'; McDowell, 'Virtue and Reason'; (Indeed, this is how Michael Smith interprets McDowell's 1978 paper: Smith, 'Humean Rationality', 77).

⁹ Nagel, *The Possibility of Altruism*, 28–30; Nagel, *The View From Nowhere*, 142; Nagel, *The Last Word*, 102–3; Millgram, 'Was Hume a Humean?', 75; Parfit, 'Reasons and Motivation', 105; Shafer-Landau, *Moral Realism*, 121; Wallace, 'How to Argue about Practical Reason', 360, 362–4; Rosati, 'Moral Motivation', sec. 3.1; Barry, 'Realism, Rational Action, and the Humean Theory of Motivation', 232; McDowell, 'The Role of Eudaimonia in Aristotle's Ethics', 21; Lenman, 'Belief, Desire and Motivation', 291; Sinhababu, 'The Humean Theory of Motivation Reformulated and Defended', 465; Wedgwood, 'Practical Reason and Desire', 346; Wong, 'Moral Reasons: Internal and External', 538; Darwall, *Impartial Reason*, 39–40.

¹⁰ In Hume's work, desires and aversions are 'passions' arising directly from sensory experiences, and hence, arising from sensory experience without any mediation from, for example, another psychological state: Hume, *A Treatise of Human Nature*, bk. 2.1.1[3], p.328.

Contemporary writers often use the term 'desire' to refer to a broader array of mental and affective states than a mere appetitive state which it may be taken to denote. In this regard, we can firstly distinguish a wide sense of a 'desire', which, as I define it here, is *any* state of motivation. If an agent voluntarily and intentionally steps toward a lamppost, then on the wide sense of desire, the agent was motivated ('desired') to take that step, however slight her motivation and for whatever justifying reason she chose to do so. According to the narrower sense of desire, as I define it, a desire state is a particular type of mental state that is not necessarily a state of motivation or being motivated. This mental state may turn out to be a necessary condition of, or part of, states of being motivated to act, but this result is not secured by definition.¹¹

It is sometimes argued that, in the context of evaluating the truth of HTM, productive discussion cannot occur if 'desire' is understood in the wide sense because this would trivialize HTM.¹² However, as Thomas Nagel and T. M. Scanlon observe, this is only relevant for theories of HTM defined as consisting solely in The Contents Principle. Where HTM is also defined as consisting in The Original Attitude Principle, use of the wider sense does not beg the question in favour of HTM because there is still a question as to whether the relevant desire is the product of a belief or process of reasoning

¹¹ The distinction between wide and narrow senses of 'desire' is drawn in different ways, although several writers define the wide sense as I do here: Locke, 'Reasons, Wants, and Causes', 172; Bond, *Reason and Value*, 11; McNaughton, *Moral Vision*, 50, 106; Schueler, *Desire*, 1, and chap. 1 generally. Indeed Schueler's 1995 text is based on this distinction; Schueler, *Reasons and Purposes*, 24; Shafer-Landau, *Moral Realism*, 139; Huemer, *Ethical Intuitionism*, 165; Skorupski, *The Domain of Reasons*, 239. And, more tentatively: Scanlon, *What We Owe to Each Other*, 37. R. M. Hare and R. Jay Wallace's definitions of a wide sense of desire may be slightly more restrictive than this: Hare, *Freedom and Reason*, 170; Wallace, 'Moral Psychology', 88–9. Some writers also define a narrower sense of desire as slightly more restrictive than that defined here. For example, Michael Huemer defines a narrower sense of desire as an appetite or emotion and John Skorupski defines a narrower sense of desire or 'want' as a non-instrumental, affective state in which I can choose not to do what I desire to do most: Huemer, *Ethical Intuitionism*, 165; Skorupski, *The Domain of Reasons*, 239.

¹² For example: Schueler, *Desire*, chap. 1.

of some kind.¹³ Nevertheless, since nothing is lost by understanding desire in the narrower sense, I adopt it here.

Secondly, writers often, but not always, use the term 'desire' more broadly to mean: **a)** any favourable attitude, and **b)** any favourable attitude irrespective of the desire's origin.¹⁴ Thus, firstly, a desire state includes, for example, states of: wanting, lusting, hankering, coveting, aspiring-to and secondly, a theory often counts such mental states as desires irrespective of whether they are derived from some further desire state or other source. I discuss the relation of desire states to similar favourable states further below. Suffice it to say here that not all mental states that resemble those on this list by being somewhat 'favourable' (such as wishes and hopes) are 'desire' states for our purposes.

Thirdly, for convenience 'desires' are often taken to include unfavourable attitudes that may explain one's motivation, such as loathing, disliking, states of disgust and repulsion as well as Hume's aversions.¹⁵ I also adopt this practice.

Fourthly, desire states need not possess a certain phenomenological quality. Just as Hutcheson recognized 'calm natural determinations of the will'¹⁶ and

¹³ Nagel, *The Possibility of Altruism*, 29; Scanlon, *What We Owe to Each Other*, 37. See Alfred Mele's and Stephen Finlay's works for a different reply, albeit one that I doubt, in the end, adequately answers the concern in question: Mele, *Motivation and Agency*, 28–9; Finlay, 'Responding to Normativity', 225–7, 227 n22.

¹⁴ Dancy, *Practical Reality*, 10–11; Smith, *The Moral Problem*, 117; Wallace, 'Moral Psychology', 88; Copp and Sobel, 'Desires, Motives, and Reasons: Scanlon's Rationalistic Moral Psychology', 254; McNaughton, *Moral Vision*, 20–1. See also: Schroeder, 'The Humean Theory of Reasons', 197–99. Cf. Shafer-Landau, *Moral Realism*, 139; Scanlon, *What We Owe to Each Other*, 37.

¹⁵ Lenman, 'Belief, Desire and Motivation', 292; Sinhababu, 'The Humean Theory of Motivation Reformulated and Defended', 470.

¹⁶ Hutcheson, 'A System of Moral Philosophy, Volume I: Concerning the Constitution of Human Nature and the Supreme Good', bk. 1.1.VI[1]–[4], pp.121–3.

Hume recognized 'calm' as well as 'violent' passions,¹⁷ contemporary writers usually acknowledge that desire states may exist despite not being felt or distinctly felt by their possessor.¹⁸ It is in light of this point in particular that I am too often misled by descriptions of the mental state in question as a 'desire'; too often I regard a desire as having a phenomenological quality. I therefore use the term 'evaluative attitude' or 'EA' instead. This term is not perfect either: writers usually think of beliefs as attitudes, and we have evaluative beliefs. However, it frees me from an implicit association with phenomenology, so I proceed with it in what follows. This means that I must distinguish EAs and beliefs, but since nothing so far said distinguishes desires from beliefs, the distinction is required in any event.

I distinguish EA states and belief states (compared to their representational content) in virtue of their different functional roles.¹⁹ In this regard it is sometimes suggested that what distinguishes belief and EA states are two particular functional roles embodying opposing directions of fit with the world.²⁰ According to this metaphor, beliefs bear a mind-to-world (hereafter, 'mind→world') direction of fit: an agent possesses a belief if the representative content of a given mental state she possesses accurately reflects a certain state of affairs in the world, at least by her own lights. In contrast, according to the metaphor, EAs have a world-to-mind (hereafter,

¹⁷ Hume, *A Treatise of Human Nature*, bk. 2.1.1[3], p.328.

¹⁸ Pettit and Smith, 'Backgrounding Desire', 568 (in regard to conscious and non-conscious desires); Smith, *The Moral Problem*, 112–3; Schueler, *Desire*, 10–11; Sinhababu, 'The Humean Theory of Motivation Reformulated and Defended', 471.

¹⁹ In light of this, I remain neutral on the issue of whether EAs and beliefs are internal states of a certain kind, mental states that are to be non-reductively understood in terms of the agent's dispositions, or states reducible to patterns of behaviour or the possibility of interpreting a given agent's behaviour as being expressive of certain EAs and beliefs. Each theory is compatible with functionalism.

²⁰ This manner of distinguishing beliefs and EAs follows the terminology writers applied to a distinction G. E. M. Anscombe developed much earlier: Anscombe, *Intention*, §32, 56; Lloyd Humberstone traces the terminology back to J. L. Austin, 1953 who used the phrase 'direction of fit' to distinguish between certain speech acts: Humberstone, 'Direction of Fit', 60.

'mind←world') direction of fit because its possessor is in favour of the world conforming to the representative content of her EA.²¹

Interpretations of the direction of fit metaphor vary according to whether they are given descriptive or normative readings and according to whether they are explicated as necessary conditions or definitions.²² For the purpose of *distinguishing* between beliefs and EAs we need only identify varying necessary conditions and these are best explicated by J. David Velleman descriptive interpretation of the metaphor. According to Velleman, for a mental state, such as a belief-that- p , to have a mind→world direction of fit is for the agent 'to be regarding the propositional object (p) as true'. In contrast, for a mental state, such as an EA-that- q , to have a mind←world direction of fit is for the agent 'to regard the propositional object (q) as to-be-made true'.

Yet this distinction is frequently conflated with a second. Since other mental states, such as assuming, imagining, hypothesizing and fantasizing, are also states in which an agent 'regards the relevant propositional object as true', Velleman observes that a belief-that- p differs from these mental states insofar as a belief also 'aims at the truth' in the sense that if one were to perceive that $\neg p$ or believe $\neg p$ (with greater credence than p), then one would cease

²¹ If there is a distinction between a state of disbelief-that- p and a state of belief-that- $\neg p$, then there is also a complementary direction of *misfit* existing for states of disbelief, doubt, skepticism. The same would be true of aversions-that- q and desires-that- $\neg q$. But I do not know of a sufficiently plausible justification for the distinction.

²² See, for example, accounts given by: Humberstone, 'Direction of Fit', 73–81; Smith, *The Moral Problem*, 115; Shafer-Landau, *Moral Realism*, 136. Dennis Stampe does not use the term 'direction of fit', but his explanation of a descriptive difference between beliefs and 'desires' could also have been described by the metaphor. Stampe, 'The Authority of Desire', 355–7.

believing that p .²³ As we can observe, the same result does not necessarily arise where one *assumes* or *imagines* that p , for example.²⁴

Correspondingly, Velleman argues that the mental states of wishing-that- q and hoping-that- q involve regarding q as to-be-made true, at least in the sense that q is not true now and there is a favourable attitude toward it being true. Yet, in contrast to these two mental states, Velleman argues that an EA-that- q ceases to exist upon one learning that q is unattainable.²⁵ He accordingly contends that the constitutive aim of an EA is the attainable. This is certainly right, but it is not the whole truth. An agent's EA-that- q may vanish for other reasons, including, for example, her learning that: **i)** her EA-that- q is based on an error; **ii)** that q is satisfied; or **iii)** that she has a stronger EA-that- r and that r is practically incompatible with realizing q .

With this understood, it is important to mention that the two characteristics of direction of fit and constitutive aim that Velleman distinguishes, need not be taken to *define* belief and EA states (it at least appears that one must also identify that what the belief state represents causes the state itself, rather than *vice versa*).²⁶ Yet, taken together, the two characteristics suffice to distinguish belief and EA states and distinguishing between the direction of fit and constitutive aims of mental states, in itself, answers certain objections to attempts to distinguish belief and EA states according to the direction of

²³ As we can see from this definition, I need not be concerned with the issue of whether belief aims at knowledge instead of truth, or the issue of whether belief also aims at some other standard in addition to knowledge or truth. On either variation, the point applies.

²⁴ Velleman, 'The Guise of the Good', 110–5; Velleman, 'On the Aim of Belief', 250; Velleman, 'Introduction', 17; Velleman, *How We Get Along*, 133–4.

²⁵ Velleman, 'The Guise of the Good', 115–7.

²⁶ In later work, Velleman further distinguishes a propositional attitude's direction of guidance. Velleman explains that the mental state of 'choice' has a direction of fit of belief and a constitutive aim of truth (just as belief does). Yet, the mental state of choice has a 'direction of guidance' similar to that of an EA state because the attitude causes what it represents, rather than *vice versa*. Velleman, 'Introduction', 25–6.

fit metaphor.²⁷ While two other counterexamples are often put to attempts to distinguish belief and EA states on the basis of this metaphor, each is answered either by iterating that this account does not purport to define EA and belief states so much as identify their relations, or by asking for the hypothetical to be spelt out in detail.²⁸ In any event, nothing highly significant depends on the particular account of beliefs and EAs I give here, so I do not answer these objections in detail here.

²⁷ For example, G. F. Schueler's objection to Michael Smith's rough explication of the direction of fit metaphor is that it entails that the state of hoping-that-*q* does not have a mind←world direction of fit, as we would ordinarily suppose, because the state would vanish upon the agent's learning that the realization of *q* is impossible. Schueler, 'Pro-Attitudes and Direction of Fit', 279–80. In a similar vein, Sergio Tenenbaum contends that this explication of the metaphor of direction of fit for beliefs is insufficiently informative because the same distinction can be made between beliefs and other attitudes, such as supposing-that-*p*. Tenenbaum, 'Direction of Fit and Motivational Cognitivism', 240. Lastly, on the basis of a particular understanding of the direction of fit metaphor, some writers suppose that EAs-that-*q* have the essential feature of disposing one to bring it about that *q* or perform a *q*-related act. As a result, literature has arisen describing counterexamples to this view. For example: Mele, *Motivation and Agency*, 25–7; Schroeder, 'Desire', sec. 1. While Mele's defence here is erroneously based on a comparison to a wish rather than an EA, ultimately we need only observe here that we can remain neutral on the issue as Velleman's explication of the metaphor is not committed to EAs having this feature.

²⁸ Sobel and Copp firstly argue that definitions of belief in terms of direction of fit must be either circular or vulnerable to counterexample because we need to refer to 'disbeliefs' to identify the conditions under which a belief goes out of existence. Sobel and Copp, 'Against Direction of Fit Accounts of Belief and Desire', 46–7. Other writers also use this counterexample in objecting to certain descriptive interpretations of the direction of fit metaphor: Coleman, 'Directions of Fit and the Humean Theory of Motivation', 130; Milliken, 'In a Fitter Direction', 564–5. While Sergio Tenenbaum also discusses this purported counterexample, he denies that it is effective against definitions of belief. Tenenbaum, 'Direction of Fit and Motivational Cognitivism', 240–3.

Secondly, Sobel and Copp contend that, just like beliefs, some EAs-that-*q* go out of existence when one comes to believe $\neg q$. Sobel and Copp, 'Against Direction of Fit Accounts of Belief and Desire', 48; Yet, I doubt that this purported counterexample works when considered in detail and, in any event, it could be explained by the EA being based on a false belief. In contrast, Mele accepts this objection and, as a result, distinguishes between 'action-desires' and 'state-desires': Mele, *Motivation and Agency*, 26. For critical discussion of this response see: Framarin, 'Motivation-Encompassing Attitudes'.

II Relative & Contingently Neutral Explanations

The problem with HTM is that it does not entail P_5 in The Relative Reasons Argument and it does not entail a conclusion incorporating an agent-relative feature (as P_2 does) that serves to disprove Moral Impartiality. To see this, let us firstly consider what is meant in P_5 by an ‘agent-relative explanatory reason’.

We defined the terms agent-relational, agent-relative and agent-neutral in regard to justifying reasons in Chapter 1 and when these terms modify reasons they have almost always, if not exclusively, been applied to justifying reasons. Of course the terms can modify explanatory reasons and other entities given that being ‘agent-relative’ just means that whether something exists or not depends on a certain standard: the individual agent or agents. The essence of the distinction then is just that, in regard to a stipulated class of rational human agents, whether a given agent could have acted in a way that could be explained by a certain explanatory reason, in certain circumstances, at a certain time depends on which agent this is.

The main point of contrast distinguishing justifying reasons in this way is that explanatory reasons are more frequently, but not always, provided at increasingly more fine-grained levels than justifying reasons. For example, an explanatory reason for an agent’s intentional act can be given: **a)** by citing facts that the agent took to be a justifying reason for acting (a ‘subjective justifying reason’);²⁹ **b)** by referencing the agent’s mental or affective states (for example, her beliefs, intentions, EAs, choices, rage); or **c)** by citing, at the microphysical level, the agent’s neurological patterns and biochemistry (which in turn may possibly be reduced to biophysics and particle physics). Other types of explanation may also be available: some academics maintain

²⁹ See Chapter 1, II for two different kinds of subjective justifying reasons.

that almost everything can be explained by economics³⁰ and some conduct may be explained by evolutionary incentives, acts of a God, and statistics.³¹

The identification of these eight types of explanations and the 'levels' at which explanations of intentional conduct can be given where reduction is admitted is not intended to be exhaustive or indicative of clear-cut levels at which further detail can be provided. Rather, the point is that in regard to the explanation of an individual rational agent's conduct, the more fine-grained the level at which the explanatory reason is given, the less likely it will be that all agents within a stipulated population will or would be able to act in a way that could be explained by precisely the same explanatory reason. Thus, for example, an agent's (*A*'s) intentional act of helping an injured girl could be explained in virtue of *A* exhibiting specific neurological pattern θ , and yet, it may not matter at all that no other agent could have intentionally helped the girl while also exhibiting θ . It may only matter that agents *B* and *C* could have intentionally helped the girl for the same justifying reason that *A* did (because she was injured and needed help). Accordingly, the explanatory reason we are interested in is that reason, the contents of which, is relevant to the agent's purpose for acting.³² The explanatory reason should then be specified in terms that are relevant and yet as ecumenical as possible, and this in two senses. Firstly, the appropriate

³⁰ For example: 'the decision to make rectangular milk cartons to minimise wasted shelf space': Frank, *The Economic Naturalist*, 21.

³¹ For further on this point see: Nikolaj Nottelmann who observes that sometimes we explain actions by pointing to the agent's occupation or social status, to the action's wider situatedness, or to the agent's emotions. As a result Nottelmann maintains: 'It appears a momentous task to argue that such explanations are always either inappropriate or reducible to some generic type of appropriate action explanation. And, anyway, as shall be clear, arguably such reductions are irrelevant to the central debates over Humeanism.' Nottelmann, 'Belief-Desire Explanation', 916.

³² Barbara Herman makes the similar point that a maxim to be tested by Kant's formula of a universal law must only consist of relevant content: Herman, *The Practice of Moral Judgment*, 74–5; A similar point about relevance is also implicit in Garrett Cullity's response to Liam Murphy's Fair Share View: Cullity, *The Moral Demands of Affluence*, 75–6.

level of explanation must be as ecumenical as possible. If it suffices to explain the agent's actions in terms of her mental states and their contents, we should not mention the agent's neurological pattern. In light of this, an explanatory reason may only need to be specified by citing the agent's subjective justifying reason for acting. In many instances, the agent's mental states, including her EAs but also her other beliefs and plans, will be relevant as well as the emotion an agent performs a certain act with, as certain virtue ethicists insist.³³ Secondly, within the appropriate level of explanation, the terms describing the reason should be relevant and yet as ecumenical as possible. If it suffices to explain the agent's conduct by citing the agent's subjective justifying reason 'that the girl was injured', then the explanation will be misleading insofar as it cites the content of the agent's subjective justifying reason as 'that the girl was injured in a leap year'. In this regard, the agent's aim in performing the act in question provides a touchstone for assessing relevance.

With this understood, we can see that if an agent's subjective justifying reason for acting is agent-relative, then the explanatory reason for that agent's motivation is also agent-relative. If an agent, *A*, has an agent-relative subjective justifying reason, *R*, to Φ in *C* at *t* then, by definition, not all other agents in the relevantly specified sub-group would also possess *R* if in *C* at *t*. As a result, not all other agents in that sub-group could Φ for subjective justifying reason *R* in *C* at *t* and the explanatory reason for *A* Φ -ing for *R* in *C* at *t* could not apply to all other agents in that sub-group.

Similarly, we can note that, at a more fine-grained level where the relevant explanation of *A*'s Φ -ing in *C* at *t* requires mentioning the influence of *A*'s

³³ For example: Aristotle, *The Nicomachean Ethics*, bk. I.viii, p.19, 1099a5–20; Hursthouse, 'Are Virtues the Proper Starting Point for Morality?', 101. Indeed, according to Michael Slote's 'agent-based' virtue ethics, right action is entirely derived from facts about the motives and dispositions of moral agents. Slote, *Morals from Motives*, 5.

mental states and their contents, the explanatory reason for A 's Φ -ing in C at t will be agent-relative in regard to a stipulated sub-group of rational agents, if this explanatory reason (including as it does now mental states and their contents) could not apply to all other agents in that stipulated sub-group, if those agents were to be in C at t . The same principle also applies in regard to reason explanations that relevantly require mentioning the agent's neurological pattern when acting. It is in this context that we can use the terms 'agent-relative belief', 'agent-relative EA' or 'agent-relative neurological pattern' in regard to a particular agent, A . In regard to beliefs and EAs, these are mental states which, relative to a stipulated sub-group of rational agents, could not be shared by all other agents in the stipulated sub-group if those other agents were in A 's circumstances, at the time in question. Similarly, agent-relative neurological patterns are neurological patterns of a given agent, A , that (similarly) could not be shared by all other agents in the stipulated sub-group, if those agents were in A 's circumstances, at the time in question.

In light of this, we can define the following three terms:

1. **Agent-Relational Explanatory Reason:** an explanatory reason that explains a certain rational human agent, A 's, intentional Φ -ing for a certain subjective justifying reason (or with a certain mental state(s) or neurological pattern(s) etc.) in C at t at an explanatory level and in terms which are relevant and as ecumenical as possible.

In contrast to justifying reasons, explanatory reasons for motivation to act are usually discussed in regard to a particular agent. They need not be so restricted: explanatory reasons can, for example, be given for the violence of a raging mob or the spending habits of certain demographics. However discussion is simplified by focussing on explanations for individual rational agents and I so limit the focus here.

2. **Agent-Relative Explanatory Reason:** an agent-relational explanatory reason explaining *A*'s intentional Φ -ing in *C* at *t* that *could not* arise for all other rational human agents in a specified sub-group that includes *A*, were those other agents, taken severally, to be in *C* at *t*.
3. **Agent-Neutral Explanatory Reason:** an agent-relational explanatory reason explaining *A*'s intentional Φ -ing in *C* at *t* that is not an agent-relative explanatory reason.

(Hence, an agent-relational explanatory reason explaining *A*'s intentional Φ -ing in *C* at *t* that *could* arise for all other rational human agents in a specified sub-group that includes *A*, were those other agents, taken severally, to be in *C* at *t*).

In virtue of this, we can see why HTM is inadequate and unnecessary in The Reason/ Motivation Argument. Even if HTM is true, so that a given rational human agent is motivated by at least an original EA and a belief properly related, this does not mean that all other rational human agents do not also possess the same original EA in virtue of, for example, their humanity or rationality. To illustrate, if it is constitutive of being a rational human agent that one has a motive-for-being-rational and if moral facts are reducible to facts about rationality, then there will be an agent-neutral relation between moral facts and moral agents.³⁴ In this case, HTM is true and Moral Impartiality is vindicated.³⁵

³⁴ For example, on the point that it is constitutive of being a rational human agent that one has a motive-for-being-rational see: Velleman, 'What Happens When Someone Acts?', 139–43; Velleman, 'The Possibility of Practical Reason'; Velleman, *How We Get Along*, 122, 127–8, 133–47. H. A. Prichard and W. D. Ross are also sometimes cited in this regard. Yet, while in his later work Prichard does argue for the existence of 'a desire to do what is right' and a 'desire to do what is a duty' (cf. his earlier 1912 paper wherein he defends a different view) and while Ross argues that the best moral motive is 'the desire to do one's duty' or what is right, neither author argues (nor for the purpose of their respective projects do they need to argue) that this 'desire' necessarily exists in moral agents. Prichard, 'Duty and Interest', 224–6; Ross, *The Right and the Good*, 158–60. For a weaker theory than Velleman's see: Schroeder, *Slaves of the Passions*, 109; Schroeder, 'Reply to Shafer-Landau, Mcpherson, and Dancy', 468–9.

³⁵ Hume maintains that the prospect of pleasure gives humans a propensity toward what gives us this satisfaction, and that the prospect of pain gives us an aversion

Furthermore, even if all rational human agents do not necessarily possess the same EA, it may still be that, necessarily, they all possess at least one of several EAs from which they could be motivated to perform the same, particular moral act.³⁶ This possibility assumes that one's motivation for acting is not relevant to what morality requires. But if this is true, an agent-neutral relation will obtain between the truth of moral propositions and moral agents. Thus, given both of these two possibilities, HTM only establishes that rational human agents always act for agent-*relational* explanatory reasons and therefore is not sufficient for **P**₅.

In addition, HTM is not *necessary* for **P**₅. **P**₅ does not require that all rational human agents are ultimately motivated by original EAs: **P**₅ is consistent with the assertion that, in some situations, an agent-relative explanatory reason is present because an agent's motivation for acting ultimately derives from her (agent-relative) *belief* that the circumstances she is presented with require her to act in a certain way in light of her own self-interest or in light of what morality requires. This is, for example, the view that H. A. Prichard and John

toward what we think generates this uneasiness in us. Hume, *A Treatise of Human Nature*, bk. 2.3.3[3], p.461. In this case it may or may not be that 'pleasure' is attained by different humans in different ways such that there is no source of pleasure that is common to all humans. Hence, it is not enough to identify a single EA or source of motivation common to all humans: it must be shown that the content of the EA in question is understood by all in a way that may allow one to claim that all humans can be motivated to Φ . But the point still remains that this may be so, in which case HTM will be consistent with Moral Impartiality.

³⁶ Richard Joyce, among others, makes this point. However he rejects the possibility that a universally applicable morality can be grounded in this way in self-interest, sympathy and rationality: Joyce, *The Myth of Morality*, 59. Mark Schroeder also makes this point and notes that it depends on the distinction between desires and reasons and therefore the falsity of what he calls the 'No Background Conditions' view: Schroeder, *Slaves of the Passions*, 109 n9. Schroeder also uses this point to explain why justifying reasons are agent-neutral, although in a much different way to the defence Joyce denies. In particular see: *Ibid.*, 97–102, 110–3, 113–7, and: Shafer-Landau, 'Three Problems for Schroeder's Hypotheticalism', 437–8; Schroeder, 'Reply to Shafer-Landau, Mcpherson, and Dancy', 469.

McDowell defend.³⁷ Alternatively, one's judgment of what she should do in a given situation may be influenced by an agent-relative worldview, such as the belief that 'money is the source of goodness' or that 'she will only gain admittance to an afterlife if she makes 2000 paper larks'.^{38/ 39} On either view, a proponent of the truth of **P₅** can agree with Nagel that a belief can generate an EA which together then explain an agent's motivation for acting.

Instead of establishing HTM then, **P₅** requires that a theory of motivation is true that establishes that all motivation (or at least all moral motivation) can be explained by an agent-relative explanatory reason. For this reason I called the theory in **P₅** 'The Theory of Relative Motivation':

The Theory of Relative Motivation ('TRM')

P₅ A rational human agent's motivation to intentionally Φ (or omit to Φ) in C at t for a sufficient or conclusive subjective justifying reason is always and only correctly explained by an agent-relative explanatory reason.

However, Moral Anti-Realists defending The Relative Reasons Argument need not stop here. They may allow that A 's motivation to act can be explained by an agent-neutral explanatory reason that is merely contingently

³⁷ Prichard, 'Does Moral Philosophy Rest on a Mistake?', 9–10 n1; McDowell, 'Are Moral Requirements Hypothetical Imperatives?'; McDowell, 'Virtue and Reason', 53–6, 69–73.

³⁸ McDowell also makes this point in terms of a given moral agent's conception of how to live, in particular, as distinct from the virtuous person's conception of how to live. McDowell, 'Virtue and Reason', 66–73.

³⁹ Michael Smith also comes to the conclusion that the real issue is not whether an agent's belief about her justifying reasons can generate a desire within her, but whether the justifying reason is itself an agent-relative justifying reason. Smith, *The Moral Problem*, 179 n3. In the context of discussing conditions on justifying reasons, Mark Schroeder also states that we need not restrict ourselves to desires: we can focus on all psychological states that have an agent-relative relation to the agent: Schroeder, 'The Humean Theory of Reasons', 196–7; Schroeder, *Slaves of the Passions*, 8–9; Finlay and Schroeder, 'Reasons for Action: Internal vs. External', sec. 1.2.3. Compare the feature that J. L. Mackie identifies as the source of the open feel of G , in Chapter 2, VII.

agent-neutral. They may allow for this insofar as proponents of a given theory of moral realism rely on an agent-neutral relation of dependence between moral facts and agents and are not content with vindicating a realist theory solely on the basis of this relation obtaining contingently. I also doubt that they would be so content. If they were, moral realism would be vindicated where, for example, in humanity's history, all explanatory reasons have been agent-relative, except at moment t where by chance the explanatory reason for A 's performing a morally relevant act in C at t could also have applied to all other moral agents in a certain class, were they in C at t . In these circumstances, a moral realist could maintain that a moral fact existed that had an agent-neutral relation to all moral agents. Yet, she must also maintain that this moral fact only existed by chance, and I doubt that realists would countenance this result, nor should they. Instead, they should maintain that moral facts are more counterfactually robust than this. There are different senses in which this may be so which we discuss below. For now, I simply note that Anti-Realists defending The Relative Reasons Argument can expand P_5 by adding the italicized disjunct to the revision of P_5 I give below. I also use this opportunity to confine P_5 to the set of the actual world and close possible worlds and to expressly identify that such Anti-Realists consider their claim about human motivation to be non-contingently true in virtue of the constitution of rational human agents. I return to this last qualification below.

Relative & Contingently Neutral Explanations ('R&CN')

P_5^* In the actual world and close possible worlds, rational human agents are so constituted that, it is non-contingently true that a rational human agent's motivation to intentionally Φ (or omit to Φ) in C at t for a sufficient or conclusive subjective justifying reason, is only correctly explained by:

- a) an agent-relative explanatory reason; *or*
- b) *a contingently agent-neutral explanatory reason.*

R&CN can be fleshed out in different ways which I need not adjudicate here. We have already briefly seen, for example, that arguably a theory of motivation need not solely consist in psychological states, but instead may provide that explanatory reasons consist in the agent's subjective justifying reason, perhaps in addition to her psychological and affective states. The agent's subjective justifying reason may consist in a fact in the world itself: 'that the café is closing is a reason to leave'. Such anti-psychologistic theories derive from the observation that we usually (but not always) try to explain an agent's motivation to act by identifying the justifying reason for which it was done or the justifying reason which the agent thought she had when she performed the act in question. For example, if one asks 'Why did Fernando aid Joan' we may reply 'because she was injured' or 'because she could have been injured'. In each case the explanatory reason given could have been Fernando's own justifying reason for his conduct and the justifying reason consists in a fact in the world and not in the agent's own psychological state(s) or a fact about such.⁴⁰ Several writers agree that facts in the world explain or partly explain agential motivation.⁴¹ However, the point is that

⁴⁰ Dancy, *Practical Reality*, 106–7. This is one of four arguments Dancy gives against what he calls 'The Three-Part Story' of action explanation: Ibid., 102–112. Dancy explains that we may take a psychological state to be a justifying reason to act or omit where, for example, one believes that 'a cliff is crumbling' or that 'pink rats are living in one's shoes'. The former belief may be a reason for one not to climb the cliff because even if the cliff is not actually crumbling, one's belief that it is may make one nervous which might not be the best state of mind to be in when climbing. The latter belief may be a reason to see a doctor: Ibid., 124–5.

⁴¹ In particular, such theories must explain how an agent can be motivated by a consideration in the world where her belief that the consideration exists is actually mistaken. Due to such cases, Russ Shafer-Landau retains a psychologistic account of action explanation. Shafer-Landau, *Moral Realism*, 178 n13. In contrast, Joseph Raz accounts for cases of mistaken belief by maintaining that, from the first-person point of view, reasons are facts, and from the explanatory point of view, beliefs and other mental states can be explanatory reasons in a different sense. Raz, 'Agency, Reason, and the Good', 23 n5; 25–6. Dancy defends his anti-psychologistic account against mistaken belief firstly by contending that the implication that the agent's reason implies a fact can be offset by adding the apposition: 'as s/he believed' to the explanatory reason; and secondly, that the transition from the first-person to the third-person point of view does not justify changing the form of the explanation from one concerned with a consideration in the world to one focused on the agent's

R&CN is consistent with either view because even if explanatory reasons partly consist in considerations in the world, the considerations an agent notices may be determined by her EAs and beliefs, as some writers maintain, and hence, explanatory reasons could still, non-contingently, be either agent-relative or contingently agent-neutral.^{42/43}

How are we to assess R&CN? While HTM is not sufficient for P_5 , it is more likely that EAs will be agent-relative compared to beliefs because of their direction of fit ('regarding so as to-be-made true'). Thus, one approach to assessing R&CN consists in firstly identifying moral and non-moral cases where an agent's motivation to act is plausibly *not* explained by an EA, or at least an original EA. Indeed, this is the approach writers take in regard to HTM. It may then be possible to argue from this point that the mental state responsible for the agent's motivation in these cases (such as a belief state) is non-contingently agent-neutral.

In the following section I consider four a priori arguments against HTM and argue that each suffers from a similar methodological problem. Given this, in

psychological states. Dancy, *Practical Reality*, 131–6. Dancy also denies the plausibility of there being two different, compatible explanations, contending firstly, that the agent's explanation cannot be the real explanation, and secondly, that this would mean that there were one too many explanations where the agent gets things right. *Ibid.*, 170–1. Cf. Nottelmann, 'Belief-Desire Explanation', 916.

⁴² Several writers maintain that our desires shape our conception of the world: Blackburn, *Ruling Passions*, 130–1, 253–6; Finlay, 'The Reasons That Matter', 17; Schroeder, *Slaves of the Passions*, 156–7. Thomas Scanlon calls these desires: 'desires in the directed attention sense'. However, he does not think that desires of this kind are always present where one is motivated to act, and, where they are present, he holds that their motivational force resides in the agent's tendency to see a consideration as a reason: Scanlon, *What We Owe to Each Other*, 39–40. In contrast, McDowell considers that an agent's conception of her circumstances is informed by her beliefs about how to live in conjunction with the particular light in which she perceives her immediate circumstances: McDowell, 'Virtue and Reason', 66–9.

⁴³ Indeed, such a theory describes what some literary writers refer to as *Blue Car Syndrome*: the experience of buying a blue car, then seeing many more blue cars on the road. The phenomenon generalizes to, for example, learning a new word. Yet, it is not so strange: it may just be that one is more inclined to *notice* those cases.

Section IV I instead approach R&CN by starting with its claim that all explanatory reasons for motivation are either agent-relative or contingently agent-neutral. From this basis, I develop an a priori argument against R&CN that also avoids the earlier methodological problem.

III The Last Word

Proponents of an Anti-Humean Theory of Motivation have sought to establish that a rational human agent's motivation to intentionally Φ does not ultimately derive from an original EA of hers, but rather from a belief, such as, a belief about possessing a conclusive justifying reason to Φ . There are several ways in which it can be argued that a belief generates motivation to intentionally Φ although two routes in particular deserve attention:

1. By directly leading to motivation to Φ itself.⁴⁴
2. By generating an EA (s), which, in turn, on its own gives rise to motivation to Φ or which, together with this belief, generates such motivation.⁴⁵

⁴⁴ Don Locke maintains that beliefs can motivate on their own. He allows that the agent also 'wanted' to perform the act, but he understands 'wanting' in the wide sense, set aside earlier, as simply meaning the agent performed the act voluntarily and intentionally. Locke, 'Reasons, Wants, and Causes', 176. S. I. Benn and G. F. Gaus can also be interpreted as maintaining this view given their contention that Nagel's 'motivated' and 'unmotivated desires' can generally be understood as forward-looking, self-regarding beliefs, and that, one can explain a person's helping the poor, merely in virtue of her subjective justifying reason to do so. Benn and Gaus, 'Practical Rationality and Commitment', 261–2, 263. H. A. Prichard also appears to defend this view in his early work where he distinguishes between acting out of a sense of duty and acting from a desire. Prichard, 'Does Moral Philosophy Rest on a Mistake?', 11–12. His later work differs: Prichard, 'Duty and Interest', 224–6; Prichard, 'Acting, Willing, Desiring', 194.

⁴⁵ Nagel, *The Possibility of Altruism*, 29–30; Nagel, *The View From Nowhere*, 151–2; Nagel, *The Last Word*, 102–111; Darwall, *Impartial Reason*, 39–40; Bond, *Reason and Value*, 11–12; Korsgaard, 'Skepticism about Practical Reason'; Stampe, 'The Authority of Desire', 348–9; Wallace, 'How to Argue about Practical Reason'; Smith, *The Moral Problem*, 148, 177–80.

In the case of each, a common Anti-Humean strategy seeks to elicit an intuition in favour of the view that a belief alone can motivate from the phenomenology of motivational experience. By 'phenomenology' I mean the philosophical method of studying the structures of one's own experiences or imagined experiences (in this case, about being motivated) from the first-person point of view. Though one is studying the structures of one's own experiences or imagined experiences from the first-person perspective, one can be doing this with a view to imagining what that experience would have been like for another person who experienced it. The range of experiences I have in mind as the object of inquiry is relatively broad, including not only perception but also thought, imagination, mental and affective states, and acts and omissions. Thus, for example, it is sometimes argued that reflection on our own experiences, from the first-person point of view, in which the content of people's thoughts or speech includes certain phrases, such as those below, provides evidence that EAs are not always required to explain one's motivation:

1. I could have stayed there for hours, but after 50 minutes my mobile phone rang. It was Nick Xerakias, my senior staff member. Malcolm had been in touch, wanting me back in the meeting. My stomach dropped. That was the last thing I wanted, but if the leader calls you, you go, no matter how you feel.⁴⁶
2. The older of the two, she is motivated by duty rather than emotion. She marries for business reasons and remains married for financial reasons. As readers, though, we begin to assign worse motives to her as the novel progresses.⁴⁷
3. I didn't want to get up. I didn't want to leave the warmth of the shelter. I didn't want to start running again in the night and in the cold.
But I had to. I had to.
I grabbed a roll off my plate and stuffed it in my pocket so I'd have something to eat later on. Then I got to my feet.
It was time to go.⁴⁸

⁴⁶ Robb, 'Mornings the Darkest Time of Day'.

⁴⁷ Cooke, *Margaret Atwood*, 147.

⁴⁸ Klavan, *The Last Thing I Remember*, 267.

As Russ Shafer-Landau observes, not only does it seem that one may be motivated by a belief in itself, but it appears that one is so motivated *despite* the presence of an aversive EA to the contrary.⁴⁹

The problem that arises for linguistic evidence in this context is that it appeals to phenomenology to elicit a favourable intuition, a strategy susceptible to a simple reply. Humeans can accept that the relevant belief(s) is partly to explain the agent's motivation, though they can assert that her conduct was ultimately initiated by the presence of an original, calm EA or an original, higher-order EA which is either calm or less noticeable in virtue of being almost always present (or a standing EA, if this notion make sense). In light of this, Humeans can explain that it is not surprising that linguistic evidence disfavours HTM because the relevant original EAs actually doing the work were unnoticeable.⁵⁰

The point has already been well made independently of linguistic evidence⁵¹ and it is widely accepted with Hutcheson and Hume that EAs need not be phenomenologically salient.⁵² Yet, writers still present arguments explicitly or implicitly appealing to phenomenology. Thus, a second Anti-Humean argument relies on an analogy to theoretical reasoning.⁵³ It is asserted that upon accepting the premises of a modus ponens argument, for example, from the proponent's own first-person experiences, she does not require an

⁴⁹ Shafer-Landau, *Moral Realism*, 123.

⁵⁰ Alternatively, as John McDowell notes, the lack of express mention of a desire in ordinary language statements about action may instead arise because the desire's presence is so obvious that it does not require mentioning, 'as when we explain someone's taking an umbrella in terms of his belief that it is unlikely to rain.' McDowell, 'Are Moral Requirements Hypothetical Imperatives?', 79.

⁵¹ Pettit and Smith, 'Backgrounding Desire', 578–80. Pettit and Smith emphasize this point in relation to what they call 'background desires' but they can just as well make the point in regard to 'foreground desires' since, as they acknowledge, both types of desire can be 'unconscious' and hence not phenomenologically salient.

⁵² See nn16–18 in this chapter.

⁵³ Scanlon, *What We Owe to Each Other*, 33–4; Korsgaard, 'Skepticism about Practical Reason', 10, 14–5; Shafer-Landau, *Moral Realism*, 124.

EA to motivate her acceptance of the conclusion. It is similar, so the argument goes, in some cases of practical reasoning: upon recognizing that I should perform my duty I do not require the EA 'to perform my duty' in order to do it. In reply to this, Thomas Nagel and James Lenman maintain that Humeans can argue that one must firstly have been motivated to accept the norm or framework governing modus ponens before one is motivated to draw the conclusion of a modus ponens argument.⁵⁴ Accordingly, one must firstly have an EA for accepting the principles governing classical logic, for example, before one draws this conclusion. Arguably such acceptance is justified (or even mandated) by a further theoretical reason or further theoretical reasons in classical logic. Perhaps it could even be said that the acceptance of the norm of modus ponens, when a rational agent reflects on it, is a condition of being rational at all, such that, her refusal to accept it demonstrates that she is irrational and that the issue is irrelevant in her case after all. Otherwise we run into the problem that the tortoise identifies.⁵⁵ But even if the acceptance of the theoretical reason justifying modus ponens upon reflection is a condition of rationality, the problem is just that in order for this theoretical reason to explain *the agent's* belief that it is true, she must have recognized and accepted this theoretical reason and it is precisely at this point that a Humean may contend that her acceptance is led by an EA. Thus, potentially lurking in the background of one's experiences on the matter is a calm, higher-order EA that explains the agent's acceptance of modus ponens as a rule and hence, the conclusion in this case that follows it. Given that this is also a plausible explanation of theoretical reasoning, without more, appeal to the phenomenology of theoretical reasoning does not suffice to show that an EA state is not present.

⁵⁴ Nagel, *The Possibility of Altruism*, 21; Lenman, 'Belief, Desire and Motivation', 296.

⁵⁵ See: Carroll, 'What the Tortoise Said to Achilles'; Blackburn, 'Practical Tortoise Raising'.

Shafer-Landau presses a third argument for the possibility that belief states alone can be sufficient to generate motivation on the basis of a mistaken self-ascription of an EA:

Suppose that you have convinced yourself of your desire to become a lawyer. Your father and grandfather were ones before you; it has always been expected of you; you find yourself in law school surrounded by others with this same desire. But you do desultory work; your attention flags as others recite cases and discuss networking over lunch; the law library enervates you. You take a summer job as a carpenter, love it, quit law school, and never look back. You reflect not long after and realize that you never really wanted to become an attorney. You thought you did; you see now that you were mistaken.⁵⁶

Shafer-Landau's point is that since the agent here discovered that she never really had an EA in favour of pursuing a career in law, she could not have been motivated to attend law school by an EA. Rather, her attending law school is better explained by her mistaken belief that she had such an EA and a corresponding means-end belief. Shafer-Landau acknowledges that speaking of the carpenter's motivation in terms of beliefs alone may be elliptical, but he maintains that since we *do* speak this way, this should be our starting point.⁵⁷ Humeanism, he claims, accordingly has the burden of disproving this view. Moreover, he defends his counterexample from the suggestion that a higher-order EA can explain her attending law school, asserting that:

[T]his seems more than a bit top-heavy. The more natural explanation simply invokes the mistaken belief (about which desire one has) and the means-end belief. These two are all that are needed to explain why the agent of our earlier example formed the intention to remain in law school, and so, in a formal sense, wanted to remain in law school.⁵⁸

⁵⁶ Shafer-Landau, *Moral Realism*, 125. Smith provides a similar illustration although merely with a view to establishing that we can be fallible about what we desire. Smith, 'The Humean Theory of Motivation', 46–7.

⁵⁷ Shafer-Landau, *Moral Realism*, 123–4.

⁵⁸ *Ibid.*, 140.

I agree that this explanation appears to be artificially moulded to answer the purported counterexample. However, the concern is readily satisfied once the content of the higher-order EA postulated is given content. Thus, in Shafer-Landau's illustration, she may have had a higher-order EA to master her chosen career or pursue a career she loves that was involved in her motivation to study law, an EA accompanied at first by a false means-end belief. This explanation is no less natural than that suggested.

Finally, Anti-Humeans contend that, irrespective of the EAs a rational human agent possesses, she can always step back and assess her situation from a detached point of view. From here she can accordingly determine what she has the most reason to do and she can then be motivated to act in this way. In contrast, Bernard Williams denies that such a perspective is achievable: when an agent steps back and reflects on her situation, her reflective self is not disassociated from her inclinations, interests, passions, fears and the like. Reflection cannot be equated with detachment.⁵⁹

This exchange recurs in Nagel's *The Last Word*. Nagel expressly acknowledges that 'Hume's position always seems a possibility' because, where prudential rationality requires 'uniformity in the weight accorded to evaluative attitudes and interests situated at different times in one's life [...] it seems possible to represent this influence as the manifestation of a systematic second-order evaluative attitude or calm passion'.⁶⁰ Nevertheless, Nagel contends that agents can always step back from such EAs and interests and ask what they should do:

The controversial but crucial point, here as everywhere in the discussion of this subject, is that the standpoint from which one assesses one's choices after this step back is not just first-personal. One is suddenly in the position of judging what one ought to do, against the background of all one's desires

⁵⁹ Williams, *Ethics and the Limits of Philosophy*, 69, 110.

⁶⁰ Nagel, *The Last Word*, 107.

and beliefs, in a way that does not merely flow from those desires and beliefs but *operates* on them—by an assessment that should enable anyone else also to see what is the right thing for you to do against that background.

It is not enough to find some higher order desires that one happens to have, to settle the matter: such desires would have to be placed among the background conditions of decision along with everything else. Rather, even in the case of a purely self-interested choice, one is seeking the right answer. One is trying to decide what, given the inner and outer circumstances, *one should do*—and that means not just what *I* should do but what *this person* should do. The same answer should be given to that question by anyone to whom the data are presented, whether or not he is in your circumstances and shares your desires. That is what gives practical reason its generality.⁶¹

Nagel acknowledges that Humeans claim that the position we stand back to is still not entirely extricated from our personal EAs, but he considers that, in this regard, HTM ‘involves a positive claim of empirical psychology’⁶² which he denies could be established on either a priori or empirical grounds.⁶³ Furthermore, he argues that, even if some of one’s reasoning is merely the expression of one’s personal inclinations, the agent can once again step back from this information and ask: ‘What, in light of all this, do I have reason to do?’⁶⁴

Despite expressly acknowledging the Humean reply, Nagel’s argument ultimately appeals to his own imagined experiences from the first-person point of view about his capacity to always be able to step back from his personal inclinations and its products. This is most clearly seen in his final claim wherein he attempts to elicit the intuition that we can always step back with detachment from our EAs and ask what we have the most reason to do. Yet, whether we can or not is precisely what is in dispute.

⁶¹ Ibid., 109–110 (emphasis present). See also: Nagel, *The View From Nowhere*, 140–2.

⁶² Nagel, *The Last Word*, 111.

⁶³ Ibid., 110.

⁶⁴ Ibid., 111.

It might be thought that Nagel's account rises above a reliance on phenomenology in one of two ways. Firstly, it might be claimed that Nagel argues in the second paragraph quoted above that in prudential reflection an agent is often seeking the (objectively) right answer. This is true; however, it does not entail that she is seeking the right answer from an entirely detached perspective, even if she thinks she is. It is always possible that it is not just her purely rational self divorced from all her EAs seeking the answer, but some so affected self. Secondly, it might be thought that, because Nagel states that the assertion that a calm or higher-order EA is present involves 'a positive claim of empirical psychology', Humeans have the burden of proving that such a state is present and at work in cases where an agent steps back from herself to reflect. However, this assertion is ultimately only justified by a phenomenological judgment, the legitimacy of which is in question.

However, even though Anti-Humeans find it difficult to disprove the possibility that a calm or higher-order EA is at work in arguments appealing to the phenomenology of motivational experience, Humeans find it equally difficult proving that the involvement of such a state *is* at work. In his discussion of Nagel's passages above, Simon Blackburn develops two main objections. Blackburn firstly observes that when agents deliberate, they do not typically deliberate about their EA states, but rather 'the relevant *features* of the external world: the cost of the alternatives, the quality of the food, the durability of the cloth, the fact that I made a promise.'⁶⁵ Blackburn avers that the agent's concerns and dispositions influence the features she attends to and how she reacts to them. But, he adds, '[t]here is not typically a *second-order* process of standing back, noticing that the cost is obsessing me [if I am a miser], and deciding to endorse that fact about myself, or alternatively

⁶⁵ Blackburn, *Ruling Passions*, 253 (emphasis present).

deciding to try and change it.’⁶⁶ The implication is that Nagel’s account of deliberation is mistaken from the outset: we do not typically even try to stand back from our EAs when deliberating.

This is right. There is not *typically* a process of standing back from one’s EAs to reflect on them. Of course, however, sometimes there is. For the very reason that agents’ concerns and dispositions do influence the features in the external world that they pay attention to and how they react to situations, agents’ sometimes at least have a justifying reason to try and withdraw from such influences and survey the scenario with ‘neutral’ eyes before considering what their best course of action is. This is often the case in situations where one senses that they are somewhat passionate about a cause or the truth of a theory. It can be advantageous to step back from one’s inclinations and survey its weaknesses with detachment (to the extent possible). Indeed, sometimes people do deliberate in this way. Nevertheless, even if they did not, this argument does not establish that we *cannot* do so.

Blackburn’s second objection is that Nagel confuses the absence of an experience of causation with the experience of the absence of causation, Arthur Schopenhauer’s distinction.⁶⁷ Blackburn thinks that we have the former (no experience of causation when we are deliberating from the impartial standpoint) and that Nagel takes this to indicate that we have the latter (an experience of being able to deliberate free from the influence of our EAs, personal inclinations and the like). But, Blackburn argues, we are not and ‘cannot be aware of all the forces that mould our agency’⁶⁸ so we cannot have knowledge of the absence of such forces from this standpoint. Accordingly, ‘we need to recognize the inevitable existence, not of a

⁶⁶ Ibid., 254 (emphasis present).

⁶⁷ Ibid., 260.

⁶⁸ Id.

perspective of free or rational agency, but of an absence that can be easily mistaken for it.⁶⁹

While Blackburn, perhaps successfully, undermines the legitimacy of Anti-Humean illustrations where agents are purportedly acting purely in virtue of a belief or reason, his argument does not go so far as to establish that EAs, interests and other personal idiosyncrasies must be influential in such situations. Blackburn establishes that we cannot be sure that we can step back from our EAs, interests and the like, as Nagel thinks we can, but Blackburn does not establish that we cannot do so at all. His claim that we 'cannot be aware of all the forces that mould our agency' cuts both ways: we cannot be aware that EAs are not influencing our agency, and, for the same reason, we cannot be aware that they always are.

The limitations of phenomenology also have implications elsewhere. Gilbert Harman argues that one cannot appeal to phenomenology to determine the truth or falsity of psychological hedonism: the hedonist can always reinterpret the data to explain why the agent in question was motivated by pleasure, whether the agent in question knew this or not.⁷⁰ A similar story is true of psychological egoism.⁷¹ Again, the problem is that there is always the

⁶⁹ Id. See also: Blackburn, 'The Majesty of Reason', sec. VI; Schroeder, *Slaves of the Passions*, 191. Sharon Street also argues that rational reflection must always proceed from some evaluative standpoint and from some evaluative premises that are taken to be true, at least provisionally, as the assessment of other evaluative judgments is undertaken. However, her only argument for this view is that it is evidenced by the widespread consensus that the method of reflective equilibrium broadly understood is our sole means of proceeding in ethics. Street, 'A Darwinian Dilemma for Realist Theories of Value', 124; Street, 'Mind-Independence Without the Mystery: Why Quasi-Realists Can't Have It Both Ways', 14–5. John McDowell and James Lenman also maintain that we cannot escape the evaluative point of view, although, without expressly defending it: McDowell, 'Values and Secondary Qualities', 119; Lenman, 'Humean Constructivism in Moral Theory', 181.

⁷⁰ Harman, 'Practical Reasoning', 457–9.

⁷¹ Harman, *The Nature of Morality*, 142. Shafer-Landau also makes this point and compares it to HTM, although with a view to concluding that evidence from

possibility that the agent's belief is subconscious or at a higher-level, thereby falling outside the parameters of phenomenology, just as calm and higher-order EAs do. Even more generally, these cases illustrate the problem of radical interpretation,⁷² and the problem of reasoning 'a priori' (even in the most permissive sense of this term) and, on this basis, drawing conclusions about one's own mental and affective states, or those of another.^{73/ 74} The difficulty is that we lack a priori justification for believing that the term we use to denote a particular internal state actually and cleanly refers to a state that the relevant agent possesses.⁷⁵

Evidence deriving from phenomenology is therefore of weak probative value and is, in itself, incapable of determining the matter, despite its contemporary use. The last word on this issue from the perspective of phenomenology is that there is no last word. Whether or not this raises a presumption against the truth of HTM depends on whether proponents of

testimony and behaviour in favour of altruism places a heavy burden on the egoist: Shafer-Landau, *Moral Realism*, 123–4.

⁷² The problem of radical interpretation is the problem that: knowing that all the physical information about an agent and her circumstances does not entail knowing her system of beliefs and desires and meanings, in her own language and ours. Lewis, 'Radical Interpretation'. In such cases it is widely accepted that the principle of charity (and perhaps also the principle of rationalization) apply so as to render people's behaviour intelligible as much as possible. However, no similar principles are available where the question in issue is what kind of mental states people have and require to be motivated at any time. Competing theories provide different answers, all of which may be equally charitable.

⁷³ One rather permissive definition of 'a priori reasoning' is: "reasoning independently of experience", where the term "experience" denotes "perceptual experience of the world beyond the thinker's body". On this definition see: "Boghossian and Peacocke, 'Introduction', 1.

⁷⁴ This problem also arises for Donald Davidson and Daniel C. Dennett who understand 'beliefs' as patterns of action and reaction. As a result both writers endorse the 'indeterminacy of belief attribution': Davidson, 'Belief and the Basis of Meaning', 153–4; Davidson, 'Appendix to Essay 10: Belief and the Basis of Meaning (1974)', 283–6; Dennett, 'Real Patterns', 46–51.

⁷⁵ Indeed it is because how we understand the world is influenced by the concepts we use to talk about it that Hilary Putnam argues that it does not make sense to talk about an absolute conception of the world 'in itself'. See: Putnam, *The Many Faces of Realism: The Paul Carus Lectures*, chaps. I–II.

Anti-HTM use it as a sword or a shield. If used as a sword to prove the truth of their own theory—to defend the point that in a particular instance, no EA at all is apparent in the agent—then the considerations identified above show that, without further evidence, the claim fails. If used as a shield, however, to defend against the claim put forward by proponents of HTM, that an original EA always lies at the source of an account of motivation, the defence succeeds unless and until proponents of this view can provide further evidence to support their claim. In the context of The Relative Reasons Argument, this conclusion therefore raises a presumption against the truth of HTM and, to a much less significant extent, the truth of R&CN. I therefore turn to a different approach to assessing R&CN.⁷⁶

IV *Morality and R&CN*

Let us firstly restate R&CN:

- P₅*** In the actual world and close possible worlds, rational human agents are so constituted that, it is non-contingently true that a rational human agent's motivation to intentionally Φ (or omit to Φ) in C at t for a sufficient or conclusive subjective justifying reason, is only correctly explained by:
- a)** an agent-relative explanatory reason; or
 - b)** a contingently agent-neutral explanatory reason.

⁷⁶ Derek Parfit reaches a similar conclusion and employs a similar strategy against normative theories of self-interest. 'Among reasons for acting, we include both moral and self-interested reasons. We can therefore ask which of these two kinds of reason is the stronger, or has more weight. As I have claimed, we may suspect that this question has no answer. We may suspect that there is no neutral scale on which these two kinds of reason can be weighed. But we do not dismiss the question as nonsensical. And we might reach an answer without finding a neutral scale. We may find arguments that can defeat the Self-interest Theory, showing that its reasons have no weight. In Part One I discussed one such argument, the claim that S is self-defeating. This argument failed. But I shall present other arguments, and I believe that at least one of these succeeds.' Parfit, *Reasons and Persons*, 129.

While ultimately P_s^* is concerned with explanatory reasons, it will simplify discussion to focus on the subjective justifying reason which the agent acts for. As identified in Section II in this chapter, if the agent's subjective justifying reason is agent-relative, then not all members of the appropriate sub-group could act for it, in which case, the explanatory reason for that agent so acting is also agent-relative.⁷⁷

This explanation also implicitly iterates the point that the terms 'agent-relative' and 'agent-neutral' are context-sensitive so that, in regard to justifying reasons, for example, an agent-neutral justifying reason is not necessarily a justifying reason for all rational agents. Rather, a justifying reason that is agent-relative in regard to all rational agents may not be agent-relative in regard to a proper sub-group, S_G , of those agents. On the contrary, it may be that a particular act, Φ -ing, in C at t performed by a certain rational agent, A , from S_G could also be performed for the same justifying reason by all other rational agents from S_G , were those agents in C at t . In such a case, the justifying reason will be *agent-neutral*, relative to S_G . Most plainly, justifying reasons of the kind we are concerned with are restricted to rational agents as a proper sub-group of the global population of human agents. By a 'rational' agent I mean an agent who **a)** can recognize and self-consciously act for justifying reasons; and **b)** has an awareness of her ordinary tendencies to respond to considerations in the world in certain ways such that, in a given situation, she could think and decide about how to respond to that situation. I take this sense of 'rationality' to preclude, for example, comatose human patients, human infants and animals.⁷⁸

⁷⁷ In Chapters 4-6 we shall also see that if the subjective justifying reason is contingently agent-neutral, then the explanatory reason is as well.

⁷⁸ In this regard I agree with Christine Korsgaard's observations on the difference between the capacity of humans and animals to reason about the world: Korsgaard, *Self-Constitution*, 115–6. Also see: Skorupski, *The Domain of Reasons*, 463–4.

Yet, as plain as it is that non-rational agents fall outside the sub-group of agents we are concerned with when determining whether or not a given justifying reason is agent-neutral, it is important to identify why this is so. While such reasons could not arise for non-rational agents at all, this is not because it would defy the laws of physics for a comatose human patient, for example, to drink a glass of juice for rehydration. It would not. Rather, it is because it is rationally impossible for her to do so in virtue of her status of being comatose. It is rationally impossible in the sense that it is impossible for her to recognize the relevant reason, know how to act for it and be motivated to do so. As a result, it is also deterministically impossible. Accordingly, what we should and do mean when specifying that a given subjective justifying reason is agent-relative or agent-neutral is that this is so with respect to *rational* agents.

The same rationale explains our focusing on *moral* (in contrast to *non-moral*, not *immoral*) agents when determining if a moral reason is agent-neutral. The conditions of moral agency are controversial but they clearly diverge from conditions of rational agency in at least one case. Young children are rational in the sense defined, yet at the ages of five and six, for instance, they lack the capacity to distinguish between moral and immoral conduct. As a result, a moral subjective justifying reason for which a moral agent Φ -ed cannot be one acted on by non-moral agents, such as young children, at all. As before, this is not because it would defy the laws of physics for non-moral agents to, for example, aid a severely injured stranger for the reason that she needs help. It would not. Rather, it is because **i)** non-moral agents rationally and deterministically cannot recognize and act for the moral justifying reason in question in virtue of their *status* as young children; and **ii)** intuitively and pre-reflectively, we consider that, nevertheless ascribing moral reasons to non-moral agents would entail that they are responsible for so acting, a pointless determination given the very formative stages of their cognitive development. As a result, what we should mean and what I

suspect we do mean when specifying that a given moral subjective justifying reason is agent-relative or agent-neutral is that this is so with respect to *moral* agents.

Consider, even more narrowly, instances in which we focus on proper sub-groups of moral agents with certain expertise. For example, imagine circumstances in which a surgeon, Dr. Hunt, makes an oversight during an operation. Call this *Surgeon*. As we know, whether or not we describe the subjective justifying reason that Hunt possibly should have had as agent-neutral or agent-relative depends not only on whether the reason is agent-relational, but also on whether it is universally possessed by members of the appropriate sub-group. Furthermore, the appropriate sub-group cannot merely be limited to moral agents at large because whether or not Hunt is morally negligent clearly should not be determined by whether her oversight fell below the standard of care expected of *all* moral agents in her circumstances. Most moral agents have little, if any, surgical knowledge and consequently have been unable to provide an appropriate standard of care at all. Instead, whether or not Hunt's oversight is negligent is determined according to a standard of care that accounts for considerations specific to the work performed by the medical profession and particularly, surgeons performing the type of operation in question. Perhaps: whether her oversight fell below the standard of care that a reasonable person with the surgeon's special knowledge and skill would be expected to provide in those circumstances. Again, this is not because it would defy the laws of physics for a surgically unqualified moral agent to take certain precautions during surgery. Rather, it is partly because it is rationally and deterministically impossible for the unqualified agent to do so at that time in virtue of his status as a surgically unqualified agent and because he was under no obligation to develop such skills and knowledge. But this is not the entire rationale. What also makes the difference in Surgeon is our chariness for

attributing moral responsibility to moral agents without the necessary expertise.⁷⁹

Compare *Ancient*. A scroll is unearthed written in Coptic and, in virtue of its potential implications certain specialists in Coptic and in ancient languages generally have a reason to translate it. Again, whether or not we describe this as an agent-neutral reason depends not only on whether the reason is agent-relational, but also on whether it is universally possessed by members of the appropriate sub-group. Given the expertise required to translate it, it is rationally and deterministically impossible for all rational agents to do this and because of this, it might be thought that the 'appropriate sub-group' only consists of rational agents with this expertise. We have seen that this rationale, in part, justifies precluding non-rational agents (generally), non-moral agents (in regard to moral reasons), and non-expert moral agents (in regard to moral reasons requiring expertise to recognize and act on). Yet *Ancient* significantly differs from each. In contrast to the rationale for precluding non-rational agents generally, non-expert rational agents are at

⁷⁹ It might be objected that an analogy between circumstances and expertise in this regard shows that expertise of any kind should not count as a feature justifying context-sensitivity. It may, for example, be argued that just as the presence of only two moral agents near a girl trying to hold onto a 50 foot ledge does not mean that only these two moral agents comprise a proper sub-group of moral agents required to help the girl, so too is it the case that, just because agents *A* and *B* have specialized expertise in neurosurgery, this does not mean that they form a proper sub-group of moral agents. Any given moral agent, *X*, would be under the same moral obligation that these two neurosurgeons are under if we superimposed on *X* the expertise that the two neurosurgeons possess. But firstly, it is not clear that all rational agents rationally could want to and could choose to become qualified experts in, for example, cryptic deciphering. Secondly, extricating specialist expertise from agents requires justification in light of the possibility of this leading to extricating agents' beliefs, evaluative attitudes, affective states and the like. If we extricate everything, all justifying and explanatory reasons would be (non-contingently) agent-neutral and this version of the distinction would have no point. Distinguishing agents' expertise from their particular physical location is one plausible point at which to draw a line for the first reason mentioned: while it is rationally possible for all moral agents to be in certain circumstances (i.e. near a ledge when someone is trying to hold on), it is not clear that all moral agents could retain a commitment to the requisite years of study to become marine biologists.

least rational agents and therefore there may be some such agents who *could have* acquired the skills and knowledge to translate ancient languages. In contrast, comatose patients and animals could not have. The distinction is therefore significant because a rational agent's not having a reason to translate the scroll due to a lack of expertise is part of what we want to know when asking whether the justifying reason to translate the scroll is agent-relative or not. Comparatively, this is not so for comatose patients; we already know this in virtue of their status.

The rationale for precluding *non-moral* agents from the appropriate sub-group in regard to moral reasons is the same as the rationale for precluding non-rational agents. The precluded agents cannot recognize and act for the justifying reason in question in virtue of their status as non-moral agents. Consequently, this rationale is also not sufficiently analogously to the purported rationale for precluding non-expert agents in Ancient.

In contrast to these two cases, in both Ancient and Surgeon there may be rational and moral agents (respectively) who could have become experts in ancient languages and surgery (respectively). In this way, the rationale in both Ancient and Surgeon differs from the rationale for precluding non-rational agents (generally) and non-moral agents (in regard to moral reasons). Yet, in contrast to Ancient, in Surgeon we are not concerned with non-expert *moral* agents because they do not possess the skills and knowledge to perform the relevant act from the outset *and*, consequently, there is no point attributing moral responsibility to them. As no such concern for moral responsibility arises for one not acting for a *non-moral* reason, such as that in Ancient, the same justification for delimiting the appropriate sub-group of rational agents to specialists of ancient languages or some other sub-group of specialists does not apply. As a result, the appropriate sub-group remains that of rational agents at large.

With this in mind, let us return to R&CN. R&CN does not merely maintain that, in the actual world and close possible worlds, it is *contingently* the case that: all explanatory reasons for all rational human agents are either agent-relative or contingently agent-neutral. Rather, just as HTM is committed to claiming that all rational human motivation in the actual world can be explained by an original EA (among other mental states), R&CN is a theory about rational human motivation in *all* circumstances and times in the actual world (as well as we can say, in certain close possible worlds). In virtue of this, in regard to rational human agents in the actual world and close possible worlds, R&CN asserts a non-contingent truth:

it is *non-contingently* true that a rational human agent's motivation to intentionally Φ is only correctly explained by either an agent-relative or *contingently* agent-neutral explanatory reason [abbreviation of R&CN].

At variance with this is a case slightly resembling Surgeon, call this *Memory*. In the future in the actual world, renowned neurosurgeons, Dr. Blake and Dr. Clarke, receive private funding to research and develop technology to diminish or remove specific unwanted memories from the minds of patients severely affected by post-traumatic stress. (If you prefer, imagine that they receive funding to research and develop technology to repair cerebrovascular systems responsible for vascular dementia, or nerves responsible for motor neurone diseases). After 5 years they succeed in developing the knowledge and skills to use certain unique technology to do so and establish a specialized practice, Group N. At present, t_1 , Group N only consists of Blake and Clarke and no other researchers are close to achieving success on this front. Hence, at the moment, only two people in the world possess the special skill and knowledge required to perform this surgery without affecting patients' other memories. Lana enters the surgery at t_1 with a high chance of attempting suicide for so long as she has a certain memory and a high chance of attempting suicide if any of her other memories are impaired.

Suppose these are circumstances in which, other things being equal, it is morally right for either Blake or Clarke to perform such surgery at t_1 and that there is a moral justifying reason for each to perform the surgery at t_1 if the other is not (and hence if no one else is). To be clear, this means that we are supposing that this justifying reason for Blake and Clarke is agent-neutral: it is agent-relational and universally possessed by members of the appropriate sub-group (namely, Blake and Clarke).⁸⁰ The supposition is legitimate because it is *imaginable* and indeed on these facts alone, probable, that this reason exists for each agent and if this is true, there is an agent-neutral moral justifying reason.⁸¹

The significant question is whether the justifying reason is *non-contingently* agent-neutral and there are different senses in which this may and may not be so.⁸² For example, even if in Memory, Blake and Clarke could have recognized and been motivated to perform the surgery for the relevant justifying reason, it also could have been the case that X did not exist, where 'X' denotes, among other things: Earth, humans, moral agents, surgeons, neurosurgeons, or Blake and Clarke. In light of these considerations, the justifying reason in question may not have existed and it might be thought that the relevant justifying reason is *contingently* agent-neutral. Yet crucially, we are not concerned with the contingency of *the justifying reason*, but rather, the contingency of *the agent-neutrality* of the justifying reason, and so, the sense in which the reason itself is contingent is not to the point.

⁸⁰ See Chapter 1, I.

⁸¹ We shall see in Chapter 4 that it is a condition for there being a justifying reason for an agent to act (or the agent *having* a reason to act—for reasons to be explained, I do not distinguish the two phrases) that she could, in a very particular sense, recognize and be motivated to act for this reason. Hence, whether the justifying reason is universally possessed or not depends on whether this is so. However, without having to explain this condition in any detail, it is again certainly *imaginable*, and perhaps probable, that Blake and Clarke could both recognize and be motivated to act on the moral justifying reason in question at t_1 . Hence, there is still a potential case of an agent-neutral moral justifying reason.

⁸² For example, see: Schroeder, *Slaves of the Passions*, 105–7.

More astutely, one could object that:

Even if we assume the existence of Group N (or some similar hypothetical case), Blake's and Clarke's justifying reasons to operate on Lana must be *contingently* agent-neutral because at t_1 Blake *could have* been: temporarily unconscious, really *not* wanting to operate, or prevented from operating for an infinite number of other explanations. Thus, Blake and Clarke's capacity to recognize and act for the relevant justifying reason at t_1 is contingent on such matters.

While this sense of contingency is certainly apparent, it is not one that we are concerned with because such events are always possible. For any moral justifying reason to Φ that we take a sub-group of moral agents to possess at t , it is always conceivable that some event may prevent one or more of these agents from Φ -ing at t , were they in the relevant circumstances. At the limit this could be a global disaster, but an agent's falling severely ill or passing out, among other events, could also detain her.

A different sense of contingency is not beholden to such events. Compare, for example, a theory about justifying reasons to act that one might regard as paradigmatically providing for non-contingently agent-neutral reasons. According to Velleman, it is constitutive of being 'an agent acting for justifying reasons' that she has a certain, singular motive that justifying reasons engage.⁸³ Whether or not Velleman is right, the theory readily provides for the non-contingency of rational agents having the requisite motive to act for justifying reasons that they recognize. If an agent does not have the motive in question, she does not fall within the sub-group of agents we are concerned with in the first place. It is because of this that, on Velleman's theory, it could not fail to be the case that an agent can act for justifying reasons and not have this motive.

⁸³ Velleman, *How We Get Along*, 120–1. Quotation marks here identify the class of agents in question; they are not used to identify a quotation. See also: Velleman, 'The Possibility of Practical Reason', 180, 199.

Furthermore, if we ask whether, on Velleman's theory, an agent's being temporarily unconscious at t_1 means that the theory only (after all) provides that it is *contingently* the case that all agents who could act for a certain justifying reason at t_1 were they in certain circumstances C, have a motive to do so, our intuitive answer is probably 'no'. At the least, in virtue of being an agent who can act for justifying reasons, the unconscious agent has the motive in question at other times, or counterfactually at t_1 . Thus, given her unconsciousness, we could:

- a) affirm that, at t_1 , she is a member of the relevant sub-group because despite being unconscious at t_1 (even if it was necessarily the case that she would be unconscious at t_1), *ordinarily* in such circumstances, she has a motive to act for the relevant reason; or
- b) affirm that, at t_1 , she is a member of the relevant sub-group where it could have been that she was *not* unconscious at t_1 , and hence that she would have a motive to act for the relevant reason; or
- c) deny that, at t_1 , she is a member of the appropriately specified sub-group of agents for whom the justifying reason was ascribed in the first place.⁸⁴

Notably, however, the sense of non-contingency expressed in **a)** is problematic in moral cases because it is not entirely consistent with the considerations rationalizing the preclusion of non-expert moral agents in cases where expertise is required, such as Surgeon and Memory. Recall that these considerations are an agent's inability to recognize and act on the reason in question, her lack of responsibility for being unable to do so and our justified chariness for nevertheless attributing moral responsibility to them. Thus, suppose that a cryptographer, Lex, ordinarily has the capacity to encode data for his government to ensure it has a sufficient degree of security and yet, at time t_x , when his government urgently needs to transmit an encoded message to save a life, Lex is unconscious after being narcotized

⁸⁴ For this third possibility also see: Schroeder, *Slaves of the Passions*, 107 n8.

an hour earlier. In the sense of non-contingency expressed in **a**), by virtue of *ordinarily* having the capacity to encode data for his government, Lex is a member of the appropriate sub-group of moral agents. But, even though Lex ordinarily could encode a message, by hypothesis, he cannot at t_x . Indeed, just as a moral agent who is not a cryptographer could not create a sufficiently secure code were he asked by Lex's government to do so at t_x , nor can Lex in his state at t_x . Lex is as blameless as the unqualified moral agent for being unable to do so. Thus, if the rationale for precluding unqualified agents from the appropriate sub-group of moral agents is an inability to act for the reason in question in C at t_x and chariness for nevertheless attributing moral responsibility to them in such circumstances, Lex should also be precluded from the appropriate sub-group. As a result, instead of grounding non-contingency in an agent's status and consequent *ordinary* capacity to perform a certain act in C at t_x , the non-contingency of agent-neutral justifying reasons is more plausibly grounded in: **1**) an agent's *actual* capacity to recognize and act for a certain reason, R, in C at t_x ; and **2**) where the agent is responsible for not being capable of acting in C at t_x for R, then her actual capacity to recognize and act for a reason to put herself in a position to be capable of doing so. (The latter conjunct serves to hold agents responsible for wilful blindness: cases where the agent wilfully prevents themselves from having the capacity to recognize certain reasons so as to avoid the potential consequences).

Shifting from moral agents' ordinary capacity to their actual capacity at a particular time t_x means that we should instead rely on the sense of non-contingency expressed in **b**) or **c**). Indeed, I rely on a conjunction of the two. Where the agent's unconsciousness or inability at t_x is not a necessary occurrence, he remains a member of the appropriate sub-group, given that it could have been the case that at t_x he could recognize and act for the reason in question (**b**). Alternatively, where his unconsciousness or inability at t_x is

in some way a necessary occurrence, then at t_x he is not a member of the appropriately specified sub-group of agents in the first place (c)).

Due to its sensitivity to agents' moral culpability, this sense of non-contingency achieves the right outcome in Lex's case as well as in Memory. In Memory it is constitutive of being a 'moral agent with the relevant special knowledge and skills at t_1 ' that one could recognize and be motivated to perform the kind of advanced neurosurgery for the relevant reason were one in C at t_1 . If a rational agent could not recognize and be motivated to do so were she in C at t_1 , then she is not a member of the appropriate sub-group of moral agents we are concerned with in the first place. In light of this, it could not fail to be the case that a member of this sub-group *could* be so motivated in C at t_1 .

Furthermore, if Blake were temporarily unconscious at t_1 this does not mean that the theory may only (after all) provide that it is *contingently* the case that all members of this sub-group could recognize and be motivated to act on the relevant justifying reason at t_1 . If her becoming unconsciousness at t_1 was not a necessary event, then if she were not unconsciousness at t_1 , she could perform the act in question in C at t_1 and she is therefore a member of the appropriately specified sub-group of moral agents for whom the justifying reason was ascribed. Alternatively, if her becoming unconscious at t_1 was an event that occurred necessarily, then we should deny that at t_1 Blake is a member this sub-group. Either way, Blake and Clarke's justifying reason for helping Lana is non-contingently agent-neutral: it is constitutive of being a member of the relevant sub-group that one could recognize and be motivated to act on a justifying reason to perform the kind of surgery in

question in C at t_1 . If one could not recognize and act on this reason in C at t_1 , then she would not be a member of the relevant proper sub-group.⁸⁵

This result is not specific to Memory. To begin with, it generalizes to all moral cases where one's justifying reason for performing a certain act in C at t arises because one belongs to a sub-group of rational agents with certain expertise.

The same is not true for *non-moral* justifying reasons requiring expertise to recognize and act on. In Ancient the appropriate sub-group in virtue of which universal possession is gauged remains the sub-group of rational agents at large. In regard to this sub-group, it could be that all actual rational agents could have a reason to translate the scroll were they in the relevant circumstances C at the relevant time t . But it is not constitutive of being a rational agent that one could recognize and be motivated to act on this reason in C at t . As a result, the justifying reason to translate the scroll in C at t is merely agent-relative or contingently agent-neutral.

The same result arises for more exoteric non-moral reasons, such as one's justifying reason to sample an orange one evening. There is no similar justification for focusing on rational agents who could recognize and act on this reason were they in the relevant circumstances on that evening.⁸⁶

Furthermore, since it is the rational impossibility of performing a certain act and the moral responsibility associated with such that underwrites our concern for a particular proper sub-group of rational agents, *all* moral

⁸⁵ According to some writers, moral reasons or facts could have a broader scope than moral responsibility. Yet, as it has been replied, there is no need for this to be so.

⁸⁶ Perhaps there are circumstances in which a sub-group of rational agents should be assigned *non-moral* responsibility to perform a certain act which justifies our confining our focus to this specific sub-group. I am yet to think of such a case that does not involve members of the sub-group entering a contract or making an agreement, but this may be a possibility.

justifying (and explanatory) reasons are non-contingently agent-neutral. The only difference between cases where moral responsibility arises due to an agent's expertise, such as Memory, and other morally relevant cases where no expertise is required is that in cases involving such expertise, it is especially *clear* that only a select sub-group of moral agents could perform the act in question in those circumstances at that time. It is especially clear because we know that developing the skills and knowledge to properly perform surgery on someone, for example, takes more than one moment in time (in June 2012). Accordingly, if an agent does not possess such expertise at a certain time t , then we know that there is nothing she can do within a certain time frame to change this. In contrast, where acting for a justifying reason, R , requires *no* special skills or knowledge, such as a reason to help an injured person, it is less clear that a particular rational agent could not recognize or be motivated to act for R . But the *degree* to which it is clear to us that an agent could not rationally Φ for the relevant reason in C at t is irrelevant: in each case she cannot rationally do so. Consequently, just as in Memory and cases involving moral reasons requiring expertise, in all moral cases the relevant proper sub-group of rational agents precludes those agents who could not rationally act for the moral reason in question in C at t .

The argument of this section then inverts a Kantian trend prevalent among writers endorsing non-contingently agent-neutral moral reasons. Such theorists maintain that, in virtue of being a 'rational' agent, one can recognize and act on moral justifying reasons.⁸⁷ In contrast, I have argued that there are non-contingently agent-neutral moral justifying reasons because from the outset, the attribution of moral justifying reasons is limited to those who: **1)** can recognize and act on them in the circumstances and at the time in question; and **2)** those who cannot, but in the past could have, recognized and acted on a justifying reason to put themselves in a position to

⁸⁷ For example: Smith, *The Moral Problem*, 148–52, 177–8.

be able to do so. Moral justifying reasons are non-contingently agent-neutral among those to whom they are properly ascribed.⁸⁸

While use of the term 'agent-neutral' is context-sensitive, its use with respect to moral agents who rationally could act for the reason in question does not merely raise a linguistic dispute. The foregoing argument does not merely maintain that we *may* choose to discuss agent-neutrality relative to these proper sub-groups; it contends that, at least in standard circumstances, we *should* do so. As a result, R&CN (P_5^*) is false.

So stated, this view may cause one to consider reasons for denying that the scope of morality should be so confined. To this end I discuss certain purported counterexamples to this view in Chapters 6 and 7 but for the purposes of this chapter, it is enough that the argument stands on its own.

VI Conclusion & Outlook

The falsity of R&CN means that The Relative Reasons Argument fails to refute all versions of Moral Realism. However, since the substance of the theory matters more than the name, it matters if the argument succeeds in denying certain versions of the theory at all as it will if P_4 and P_1 are true. Indeed, I think the truth of these two premises is plausible, and in Chapters 4-7 I defend a particular version of each via what is both the only available route, and one that is, so far as I know, yet to be defended.

⁸⁸ It is also for this reason that the issue of whether agent-neutral justifying reasons have explanatory priority over agent-relational reasons, or *vice versa*, is less important. On either view, we shall only be concerned with the same sub-group of moral agents.

4

Justifying Reasons & the Possibility of Motivation

"Which is it? You must have come here for some reason."

"Well, I—" Milo began.

"Come now, if you don't have a reason, you must at least have an explanation or certainly an excuse," interrupted the gateman.

Milo shook his head.

"Very serious, very serious," the gateman said, shaking his head also. "You can't get in without a reason." He thought for a moment and then continued. "Wait a minute; maybe I have an old one you can use."

He took a battered suitcase from the gatehouse and began to rummage busily through it, mumbling to himself, "No . . . no . . . no . . . this won't do . . . no . . . h-m-m-m . . . ah, this is fine," he cried triumphantly, holding up a small medallion on a chain. He dusted it off, and engraved on one side were the words "WHY NOT?"

—Norton Juster, *The Phantom Tollbooth*, chap. 3

Chapter 3 started considering The Relative Reasons Argument and concluded that P_5^* , Relative & Contingently Neutral Explanations, is not always true. If this is right, then The Relative Reasons Argument is unsound. Nevertheless, if P_1 and P_4 are true, then significantly, if there are moral facts, they are dependent on the capacity of moral agents to recognize and be motivated to act in light of them. I think this is the case and in the present chapter and Chapters 5 and 6 I defend the truth of a particular version of P_4 while in Chapter 7 I defend the truth of P_1 .

P₄ of The Relative Reasons Argument states:

P₄ All sufficient and conclusive practical justifying reasons for a rational human agent, *A*, to Φ in *C* at *t* are necessarily related to *A*'s capacity to be motivated to intentionally Φ (or omit to Φ) in *C* at *t*. (*Reason/ Motivation Internalism*)

The opposing theory is:

Reason/ Motivation Externalism: \neg (Reason/ Motivation Internalism).

The necessary relation in Reason/ Motivation Internalism may be spelt out in several ways. For example, the relation may be **1.** actual or counterfactual; and may **2.** relate to an actual state of (at least) some motivation or a mental or affective state (such as an evaluative attitude ('desire'), desire, belief, intention etc.) that plays a necessary role in an agent having such an actual state of motivation.¹ Indeed, there need not be only *one* necessary relation that obtains. Just as for evaluative attitudes and motivation in Chapter 3, all that is required for The Relative Reasons Argument is that: necessarily, there is a relation between a given justifying reason and the relevant agent's motivational capacity, whether or not it is the same relation in each instance. **P₄** could be so qualified.

Of these variations, the most oft-discussed is defended by Bernard Williams and I outline his version (hereafter: 'WIR') in Sections I and II below.² Since I dispute WIR and defend the existence of a different necessary relation, I am not so concerned with interpretative issues in these sections. However, explaining certain aspects of WIR is crucial for delineating the version of

¹ This final possibility is noted by: Audi, 'Moral Judgement and Reasons for Action', 154; Finlay and Schroeder, 'Reasons for Action: Internal vs. External', sec. 1.1.1.

² Williams initially sets out and defends his view in: Williams, 'Internal and External Reasons'.

Reason/ Motivation Internalism I defend, so discussion of certain aspects of WIR is necessary. Finally, Section III explains why WIR is vulnerable to the conditional fallacy and outlines the version I defend, what I call *Modest Internalism*.

For all this, discussion of any kind of relation between justifying reasons and motivation may appear presumptive because the term ‘justifying reason’, its referent, and concept are the subject of 16 or more issues in the literature. However, ultimately these issues are orthogonal to whether there is a relation between reasons and motivational capacity. Several such issues are themselves concerned with a relation between justifying reasons and something else (e.g. values, morality, rationality, other senses or kinds of reasons), others have no bearing on this matter (e.g. the ontological and temporal statuses of reasons, holism about reasons, cognitivism about reasons), and issues revolving around the analysis of the concept of a justifying reason in the normative or non-normative domain must nevertheless account for this relation obtaining or not obtaining. That is, the issue of whether justifying reasons are necessarily related to agents’ motivational capacities is itself relevant to analyses of justifying reasons.³

³ Joshua Gert is the only author I know who makes use of general characteristics of justifying reasons in taking a stance on this issue. Gert distinguishes between justifying reasons that give one *permission* to act (which he maintains can be external reasons) and justifying reasons that *require* one to act (which he maintains are all internal). Ultimately, the basis for his distinction is reflected in the distinction I shall draw between pro tanto justifying reasons on the one hand, and sufficient and conclusive justifying reasons on the other hand. Gert, ‘Skepticism about Practical Reasons Internalism’, 69; Gert, ‘Internalism and Different Kinds of Reasons’, 55–6; Gert, ‘A Functional Role Analysis of Reasons’, 356–9.

I Williams' Internalism

According to Williams, statements of the form '*A* has a reason to Φ ' and '*There is a reason for A to Φ* ' must meet a necessary condition to be true. His favoured formulation of the condition is:

A has a reason to ϕ only if there is a sound deliberative route from *A*'s subjective motivational set [per Williams' abbreviation: '*S*'] to *A*'s ϕ -ing.⁴

In this formulation, ' ϕ ' stands for some verb of action and 'a reason' refers to both pro tanto and conclusive reasons.^{5/ 6} While Williams thinks that this condition must be met for the above statements to be true, he believes that they are asserted (and interpreted) in two different senses:

1. a sense that is intended to comply with this necessary condition (an 'internal reason statement'); and
2. a sense that is not intended to comply with this necessary condition (an 'external reason statement').

⁴ Williams, 'Postscript: Some Further Notes on Internal and External Reasons', 91[2] (emphasis present); Williams, 'Replies', 186[3]. Williams 2001 formulation differs in two respects from his 1989 formulation: **1.** In his 2001 formulation there is no requirement that the agent himself has to be able to reach the conclusion to Φ by the sound deliberate route from *S*; rather now it only has to be the case that there is a sound deliberative route from *S*; and **2.** The sound deliberative route no longer must terminate in a conclusion to Φ ; rather now it must reach further to *A*'s actually Φ -ing.

⁵ This is demonstrated in Williams' comments in 2001: 'It is natural to take the condition [Williams' necessary condition] as implying not just that *A* has a reason to ϕ , but that he or she has more reason to do that than to do anything else. This is the case I shall take as central.' Williams, 'Postscript: Some Further Notes on Internal and External Reasons', 91[2]. See also: Williams, 'Internal Reasons and the Obscurity of Blame', 39[1].

⁶ cf. Pettit and Smith, 'External Reasons', 143–4 where the authors consider that Williams is only concerned with pro tanto reasons citing: Williams, 'Internal and External Reasons', 104[2]. I disagree however that this passage excludes Williams' concern for conclusive reasons.

Williams' contention is accordingly that, despite their use in ordinary conversation, all external reason statements are either false or they actually express something else; only internal reason statements can be true. The corollary is that there are no external reasons; there are only internal reasons.

Williams understands the content of the necessary condition in a particular way. He non-exhaustively defines 'S' as the agent's 'desires' as well as her 'dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent'⁷ and her 'evaluations, attitudes, projects, and so on'.⁸ While he maintains that there is no attempt in his account to exclude altruistic or ethical considerations,⁹ he does exclude the agent's 'needs'.¹⁰ There is some disagreement about whether Williams' S therefore consists of 'desire-like' states (in the sense, let us say, of evaluative attitude states defined in Chapter 3), or any motivational state. However, since he clearly eschews a commitment to The Humean Theory of Motivation, as we shall see, I follow the latter interpretation here.¹¹

Williams also non-exhaustively defines 'a sound deliberative route', only explaining that it involves 'at least correcting any errors of fact and reasoning involved in the agent's view of the matter'.¹² He expressly cites

⁷ Williams, 'Internal and External Reasons', 105[2].

⁸ Williams, 'Internal Reasons and the Obscurity of Blame', 35[2].

⁹ Williams, 'Internal and External Reasons', 105[2]; Williams, 'Postscript: Some Further Notes on Internal and External Reasons', 92[2].

¹⁰ Williams, 'Internal and External Reasons', 105[3]–6[1]. From this description it is unclear whether Williams understands S to consist in mere desire-like states or something more expansive such as dispositions to respond to recognized reasons. For an excellent discussion of this see: Finlay, 'The Obscurity of Internal Reasons', 3–13.

¹¹ This interpretation is also objected to by Stephen Finlay and while I disagree with these objections, what I want to say does not depend on correctly interpreting Williams on this point. I therefore leave this issue aside. Finlay, 'The Obscurity of Internal Reasons', 9 col. 2 – 10 col. 1.

¹² Williams, 'Internal Reasons and the Obscurity of Blame', 36[3].

correcting errors of fact and reasoning because firstly, he explains, it is very odd to say that 'A has a reason to drink a glass of petrol' when A believes it to be a glass of gin. It is more natural to say 'he does not have this reason, though he thinks he has'.¹³ Secondly, he argues that internal reasons are not solely concerned with explanation, but also the agent's 'rationality'¹⁴ and that 'any rational deliberative agent has in his S a general interest in being factually and rationally correctly informed'¹⁵ and 'to take the correct means to his ends'.¹⁶

According to Williams this means that whether an agent has, in the end, a reason to act is *not* determined by her pre-existing S. Rather, it is determined by her S as this would be when she has no relevant false beliefs and all relevant true beliefs of the situation at hand, including beliefs about the contents of her S. To keep this in mind I let 'S_B' denote A's S after these corrections to her beliefs.^{17/18}

¹³ Williams, 'Internal and External Reasons', 102[4]. Quotation marks here clarify the content of what it is purportedly natural to say rather than identifying quotations.

¹⁴ Ibid., 102[5]–3[1].

¹⁵ Williams, 'Internal Reasons and the Obscurity of Blame', 37[2].

¹⁶ Williams, 'Postscript: Some Further Notes on Internal and External Reasons', 92[1].

¹⁷ It may appear somewhat artificial to break A's process of idealization into stages in this way but this is merely invoked to aid understanding. In any event, appealing to a more idealized version of A to determine what reasons A has is itself an artificial device.

¹⁸ Despite this, Williams' understanding of 'sound deliberation' does not require that an agent be concerned for her long-term prudential interests or for moral considerations. See his argument for this and qualification of long-term prudential interests at: Williams, 'Internal and External Reasons', 105[4]–106[1]; Williams, 'Internal Reasons and the Obscurity of Blame', 37[2]; Williams, 'Postscript: Some Further Notes on Internal and External Reasons', 92[1]. In contrast, if everyone has an interest in having no relevant false beliefs, all relevant true beliefs, and correct reasoning, as Williams suggests, then I think it will turn out that everyone has long-term prudential interests. That morality can also be grounded in rationality is, however, controversial and we discuss this briefly in Chapter 7, II.

A. Kinds of Reasoning

In addition to correcting errors of fact, Williams understands sound deliberation as correcting errors in reasoning. According to Williams, the kinds of reasoning we are concerned with include Hume's theory of practical reasoning, wherein, on some interpretations, reasoning merely serves to determine the most efficacious means to realizing or constituting one's ends.¹⁹ However, Williams also lists other forms of reasoning, such as resolving conflicts between elements of one's S_B ,²⁰ and explains that other forms may obtain because thinking about what to do requires, *inter alia*, imagination and it is impossible to anticipate how this could influence deliberation.^{21/ 22}

Williams also considers that deliberation can result in new *actions* for which one has internal reasons²³ and he allows that, contrary to Hume, an agent's deliberation can add and subtract elements from her S and S_B .²⁴ Yet there is disagreement about whether Williams intends to take A 's S_B as it is or after A correctly deliberates. While I am more concerned about substantive issues, there is sufficient evidence for the latter view given, as we have just seen,

¹⁹ The main alternative interpretation argues that Hume was skeptical of there being practical reasoning and/ or practical reasons at all. See for example: Millgram, 'Was Hume a Humean?'; Korsgaard, 'The Normativity of Instrumental Reason', 222, 222–3 n23.

²⁰ The other forms are: determining how to satisfy multiple elements of one's S_B by, for example, time-ordering; finding a specific form for a project adopted in unspecific terms; inventing alternative means and ends; and allowing for the perception of unexpected similarities: Williams, 'Internal and External Reasons', 104[3]; Williams, 'Internal Reasons and the Obscurity of Blame', 38[2].

²¹ Williams, 'Internal Reasons and the Obscurity of Blame', 38[2].

²² I assume that Williams does not think that 'practical reasoning' concludes in action so that 'reasoning without error' actually means a process that concludes in A Φ -ing.

²³ Williams, 'Internal and External Reasons', 104[4].

²⁴ *Ibid.*, 104[4], 108[3].

that Williams expressly emphasizes that deliberation may alter A 's S .²⁵ In contrast to this, John Skorupski argues that Williams frequently refers to ' A 's existing motivations' or ' A 's motivations A already has', where emphasis is added to the modifiers 'existing' and 'already'.²⁶ However, *pace* Skorupski, I doubt that these qualifications are intended to emphasize a contrast between A 's S_B and ' A 's motivations after deliberation.' If the addition of 'existing' was intended to emphasize A 's motivations *prior to correction*, this explanation of the modifier would just as suitably suggest interpreting ' A 's existing motivations' as A 's \underline{S} (compared to A 's S_B). Yet since we know that Williams intends A 's S to at least be taken as A 's S_B , this explanation does not seem right. Furthermore, it is natural to understand ' A 's existing motivations' as ' A 's motivations after idealization' in contrast to 'the motivations A might have at any later time after deliberation' because without the modifier 'existing', the reference is ' A 's motivations', a description that does nothing to deter the interpretation ' A 's motivations at any time at all'.

I therefore distinguish ' S_{BR} ' as a third version of S arising when S_B is revised from further reasoning. This process can, at least theoretically, continue to some extent. A can subsequently reason from her $\underline{S_{BR}}$ and this may result in further changes to her S . I therefore let ' S_{BR+} ' denote a state of A 's S arising subsequent to her correctly reasoning either from her S_{BR} or her S_{BR+} . It denotes ' a state of A 's S ' because, as discussed above, A may have more than

²⁵ 'Reflection may lead the agent to see that some belief is false, and hence to realise that he has in fact no reason to do something he thought he had reason to do. More subtly, he may think he has reason to promote some development because he has not exercised his imagination enough about what it would be like if it came about. In his unaided deliberative reason, or encouraged by the persuasions of others, he may come to have some more concrete sense of what would be involved, and lose his desire for it, just as, positively, the imagination can create new possibilities and new desires. [...] *We should not, then, think of S as statically given. The processes of deliberation can have all sorts of effect on S* ', 104[4]–5[2] (emphasis added).

²⁶ John Skorupski, for example, puts forward this view: Skorupski, 'Internal Reasons and the Scope of Blame', 78–9.

one potential correct process of reasoning, and hence, in a given case, there may be more than one resulting S_{BR+} . In light of this, Rachel Cohon objects that without a clearer understanding of ‘correct process of reasoning’ we cannot identify whether rational agents’ conclusions *are* actually restricted by their motivational set.²⁷ Yet, it suffices that both parties agree that there are certain cases in which A could not rationally recognize or be motivated to Φ for a given justifying reason. Therefore, for convenience, when discussing the agent in question, I let ‘ A_{BR+} ’ denote ‘ A with S_{BR+} ’.

B. Cognitivism about Reasons

Significantly, Williams requires not just that there is a sound deliberative route from an agent’s S to her Φ -ing, but to her Φ -ing *for that reason*.²⁸

If it is true that A has a reason to ϕ , then it must be possible that he should ϕ for that reason; and if he does act for that reason, then that reason will be the explanation of his acting.²⁹

He also implies that A must *consciously* Φ ‘for that reason’.³⁰ Accordingly, hereafter in using the phrase *for that reason* in the context of ‘an agent’s reason to Φ for that reason’ I do not merely mean: **a)** that A Φ ’s consistently with this justifying reason whether or not she is aware of either the consideration giving rise to the reason to Φ or the reason to Φ itself; or **b)** that A Φ ’s consistently with this justifying reason and in awareness of, or

²⁷ Cohon, ‘Are External Reasons Impossible?’, 555.

²⁸ Prior to Williams, Don Locke defended this view: Locke, ‘Reasons, Wants, and Causes’, 170.

²⁹ Williams, ‘Internal Reasons and the Obscurity of Blame’, 39[1]; Williams, ‘Internal and External Reasons’, 102[4], 106[4]; Williams, ‘Postscript: Some Further Notes on Internal and External Reasons’, 93[1].

³⁰ ‘The basic case must be that in which A ϕ ’s, not because he believes only that there is some reason or other for him to ϕ , but because he believes of some determinate consideration that it constitutes a reason for him to ϕ .’ Williams, ‘Internal and External Reasons’, 107[2]. (Williams italicized variables in his original paper, but not in his second and third paper on the issue thus, italics here are his).

because of, *the consideration* that gives rise to this justifying reason, whether or not she is aware of the *justifying reason* to Φ . Rather I mean:

for that reason: that an agent Φ 's consistently with this justifying reason because she recognizes *the consideration*, that gives rise to the justifying reason for her to Φ , as giving rise to that justifying reason for her to Φ .³¹

Call this feature of Williams' necessary condition: Cognitivism about Reasons.³²

Reason/ Motivation Internalism in conjunction with Cognitivism about Reasons requires that the agent be able to: 1) 'see' or recognize the reason; and 2) be motivated to act for that reason. As these are separate acts, proof of the impossibility of either refutes Reason/ Motivation Internalism when taken in conjunction with Cognitivism about Reasons (and hence it will refute WIR). In this regard, 'possibility' or 'impossibility' may be understood in different senses.

C. Possibility

We already defined a sense in which it is *rationally* possible for a rational agent to be motivated to perform a certain act in Chapter 3: she can recognize the relevant reason to do so, know how to act for it and be motivated to do so. This is also the sense of possibility I shall be concerned with here and one other writers defend, albeit with slightly different definitions.³³

³¹ This point takes into account that sometimes the best way for one to realize a certain result is by pursuing it obliquely. I discuss this further below (Chapter 4, III).

³² This title is used in different ways. Compare uses given by: Bratman, 'Cognitivism about Practical Reason', 117; Setiya, 'Cognitivism about Instrumental Reason', 650–1; Ross, 'How to Be a Cognitivist about Practical Reason'.

³³ Williams, 'Internal and External Reasons', 103[1]–[2], 103[5]; Darwall, *Impartial Reason*, 41–2; Korsgaard, 'Skepticism about Practical Reason'; Shafer-Landau, *Moral*

Other senses of possibility are problematic in this context. As Russ Shafer-Landau remarks, *logical* possibility is too weak because a reason may logically follow from any motivational fact. He also states that *physical* possibility is too weak and too strong. It is too weak, he discerns, because it may be physically possible that anyone can believe or possess an evaluative attitude about anything and it is too strong if determinism is true and only one course of action is in fact physically possible.

Furthermore, it is not clear what: 'it is *metaphysically* possible for *A* to be motivated to Φ in *C*' means unless this is interpreted as:

the event of '*A* being motivated to Φ in *C*' could occur in the actual world, if not other possible worlds.³⁴

But what does 'could' mean in this interpretation? As the event '*A* being motivated to Φ in *C*' includes a rational agent being motivated to act, it must at least mean that it is physically possible for *A* to Φ in *C* in the sense that the event does not defy the laws of physics and it does not appear to mean more than this. So, recourse to the metaphysical sense is unhelpful.

Cognitive possibility, in contrast, is the capacity to recognize a certain justifying reason and the capacity to know how to act on it. This is part of what we are concerned about, where the question of 'whether *A* could know that there is a certain reason for her to Φ in *C* at *t* after deliberating

Realism, 171–2; Smith, *The Moral Problem*; Smith, 'Internal Reasons'. Pettit and Smith have been interpreted as defending WER in: Pettit and Smith, 'External Reasons'. Yet, it is unclear from this paper whether Pettit and Smith are merely criticizing WIR or whether they are, in so criticizing Williams, intending to prove the truth of WER generally.

³⁴ See, most recently: Nolan, 'The Extent of Metaphysical Necessity' where the author distinguishes several senses in which something may be 'metaphysically necessary' in virtue of the range of the domain of worlds the thing applies to.

rationality', extends to *A*'s knowledge and circumstances after having taken further investigative steps, where it is relevant to do so.

Yet, since this is not just a question about what *A* could *know*, but about whether *A* could be in a certain motivational state, even upon recognizing that there is a certain justifying reason for her to Φ in *C* at *t*, rational possibility in full is required. *A* must also have the capacity to be motivated to Φ in *C* at *t* for a given justifying reason that she *already* recognizes (*motivational* possibility). Rational possibility therefore consists in the conjunction of cognitive and motivational possibility. Specifically, that it is 'rationally possible for *A* be motivated to Φ in *C* at *t*', means that it is possible for *A* (or a more ideal version of *A*) to recognize the relevant reason, *R*, to Φ in *C* at *t*, know how to Φ in *C* at *t* for *R*, and be motivated to Φ in *C* at *t* for *R*.³⁵ As a corollary, Milo may not always be able to Φ for the subjective reason 'why not Φ ?': he must firstly recognize the reason and his countervailing reasons not to Φ in a given case, may prevent him from being able to be motivated to Φ for this reason.

D. Weak, Ignorant, Akratic & Strong Internalism

Distinguishing WIR's requirements of an agent recognizing a justifying reason and being motivated to act for it entails four versions of counterfactual internalism:^{36/ 37}

³⁵ The adjective 'rational' is then being used in this context in a relatively minimal sense and one I distinguish from the use of the term 'rational' and its cognates in contexts in philosophy not limited to discussion of WIR. This includes, for example, Stephen Finlay's use of the term in a sense in which 'what it is rational for a certain agent to do' is not restricted by the agent's original *S*. Finlay, 'The Obscurity of Internal Reasons', 3 col. 1, 8 cols. 1–2, 8 n25.

³⁶ Williams' papers are silent on this issue. There are hints that he adheres to Weak Internalism in: Williams, 'Internal and External Reasons', 106[4], 110[3]; Williams, 'Internal Reasons and the Obscurity of Blame', 35[2], 39[1]; and there are hints he is more willing to adhere to Strong Internalism in: Williams, 'Replies', 189[3], 190[2], 191[2]; and in: Williams, 'Postscript: Some Further Notes on Internal and External

It is a necessary condition of 'A having a reason to Φ ' (or 'there being a reason for A to Φ ') that:

- Weak Internalism:** A_{BR+} *could* recognize the reason in question and, upon doing so, *could* act for that reason.
- Ignorant Internalism:** A_{BR+} *could* recognize the reason in question and, upon doing so, *would* act for that reason.
- Akratic Internalism:** A_{BR+} *would* recognize the reason in question and, upon doing so, *could* act for that reason.
- Strong Internalism:** A_{BR+} *would* recognize the reason in question and, upon doing so, *would* act for that reason.

Both Strong and Akratic Internalism are implausible. That A_{BR+} *would* recognize a reason to Φ is perhaps plausible in relatively simple cases, such as whether an agent has a reason to sip the champagne overdosed with an hallucinogen. A_{BR+} 'would' (we are inclined to say) *not* sip the champagne, other things equal. But the *ceteris paribus* clause is telling because it functions to ensure the truth that A_{BR+} *would* not sip the champagne. Without it we cannot be so sure that this would be the case because there might be some scenario in which A_{BR+} does have that reason. Thus, the assertion that A_{BR+} would definitely recognize a certain reason in a given case is falsified by the recognition that A_{BR+} 's imagination may play a greater role in reasoning than

Reasons', 92[2] insofar as he draws an analogy to advising someone in the 'if I were you...' mode. In this regard one does not tend to say 'if I were you I could Φ ' rather one says 'if I were you I *would* Φ '. But his work does not conclusively answer the issue. Perhaps this is because he leaves it unanswered what S consists in, what facts contribute to sound deliberation, and what kinds of reasoning are employed in sound deliberation. Or, perhaps he did not consider the variable.

³⁷ Derek Parfit distinguishes between Weak and Strong Internalism but not Akratic and Ignorant Internalism. I am following Parfit in using the former appellations: Parfit, 'Reasons and Motivation', 117. Kent Hurtig also distinguishes between 'actual' and 'possible' variations of Internalism but the versions of possible internalism he criticizes do not include Weak and Ignorant Internalism. Hurtig, 'Internalism and Accidie'.

Akratic and Strong Internalism allow. Accordingly, there may be more than one way A_{BR+} could reason without error, particularly in complicated, non-moral cases. Williams implicitly acknowledges this:

there is an essential indeterminacy in what can be counted a rational deliberative process. Practical reasoning is a heuristic process, and an imaginative one, and there are no fixed boundaries on the continuum from rational thought to inspiration and conversion.³⁸

This passage does not explicitly say that there can be more than one way to reason in some situations. Rather it says that what is to count as a process of practical reasoning in the first instance is left indeterminate. However, the latter entails the former. If what is to count as a process of practical reasoning is left indeterminate and if the particular states that comprise S is left indeterminate, in some cases there will be more than one reason for acting that A_{BR+} could determine without error. If this is right, then we cannot say what A_{BR+} would certainly and would not certainly recognize as a reason to act in a given case.

Ignorant Internalism is also problematic. It has the burden of explaining the impossibility of A_{BR+} suffering *akrasia*, a burden that I doubt it can fulfil. That A_{BR+} has all relevant correct information and reasons without error will assist A_{BR+} to see that she has a conclusive reason to Φ , but it does nothing to ensure that A_{BR+} is actually motivated to act in accordance with the reason recognized.^{39/ 40}

³⁸ Williams, 'Internal and External Reasons', 110[3].

³⁹ Of course, in the context of practical reasoning, some writers would understand 'reasoning without error' as reasoning that issues in the agent actually acting as she concludes that she should. This view supposedly originates in Aristotle's work and is recently taken up by Sergio Tenenbaum and Christine Korsgaard. Tenenbaum, 'The Conclusion of Practical Reason'; Korsgaard, *Self-Constitution*, 124–5. I remain neutral on this point. If action is taken to be the conclusion of practical reasoning then this may very well insulate A_{BR+} from *akrasia* and this may mean that A_{BR+} would Φ in C . However this view does not relevantly differ from the present position which also maintains that there are instances in which A_{BR+} does not suffer from

Since Strong, Akratic and Ignorant Internalism are implausible, one could interpret Williams as adhering to Weak Internalism. In Weak Internalism, the first instance of 'could' denotes cognitive possibility, and the latter, motivational possibility. It may be thought that each instance of 'could' (instead of 'would') is problematic because it leaves it unclear when A_{BR+} could and when she could not recognize and Φ for the reason in question. Yet firstly, as we have seen, such vagueness derives from Williams' incomplete definitions of S and the types of reasoning available to A_{BR+} and Williams celebrates the vagueness of these two definitions as a mark of authenticity.⁴¹ Furthermore, notwithstanding this, vagueness does not vitiate the existence of the necessary condition, it just means that its fulfilment in a given case may be difficult to determine or indeterminable. Thirdly, the vagueness of the necessary condition does not render it impervious to criticism. There are cases in which Weak Internalists may clearly deny that the condition is fulfilled. For example, in 1892 King Leopold II forced Congolese men into slave labour, cutting off the hands of those who failed to produce a quota of rubber, murdering their children and burning their village. Weak Internalists may agree that, at a particular point in time in this period, it was rationally impossible for Leopold (with S_{BR+}) to recognize a reason to be merciful to a given Congolese worker, or that there was something in his S_{BR+} allowing him to be so merciful.

akrasia and would Φ in C. As Pettit and Smith mention, some writers may also think that the conclusion of practical reason is a statement in a different mood such as: 'Let me Φ '; 'I will Φ ' or ' Φ -ing it is' Pettit and Smith, 'Backgrounding Desire', 571.

⁴⁰ Kent Hurtig mentions the possibility of qualifying the internal relation such that it applies absent *accidie* and denies this view. Hurtig argues that this violates William's requirement that deliberation must proceed from the actual motivations of the agent. However, Hurtig's argument is unsound: even though the deliberation must proceed from the agent's actual motivations this does not mean that the agent *must* be *akratic*. Hurtig, 'Internalism and Accidie', 532–3.

⁴¹ 'It is often vague what one has a reason to do. For one thing, the lines between rationality and imagination are vague.' Williams, 'Internal Reasons and the Obscurity of Blame', 38[2] (emphasis present).

Finally, we can stipulate that A_{BR+} would have to be able to avoid the problems that Strong, Akritic and Ignorant Internalism encounter before a Weak Internalist agreed that A_{BR+} could rationally Φ in the case in question. Thus, for example, if A_{BR+} would *necessarily* suffer from *akrasia* in a given case then Weak Internalism would not maintain that A_{BR+} *could* rationally Φ in C for the reason in question.⁴²

This still leaves unanswered the question of whether each sense of ‘could’ in Weak Internalism is relativized to: 1) the particular circumstances, C, and time, t , in question; or 2) A ’s capacity to recognize and be motivated to act for the relevant justifying reason *generally* or *in most circumstances and at most times*, even if not in C at t . Consistently with the sense of contingency adopted in Chapter 3, the version of Reason/ Motivation Internalism I defend adheres to the former view because, as we discuss in Chapter 5, I see no need to revert to the latter view, even though it countenances a more expansive set of internal justifying reasons.

II Externalism

Williams defines ‘externalism’ about reasons (hereafter ‘WER’) as the negation of WIR:⁴³

⁴² In this way, Weak Internalism avoids the charge that Russ Shafer-Landau and Mark Schroeder put to it, ‘that it does not allow for *akrasia*’ and that Kent Hurtig puts to ‘actual internalism’, ‘that it does not allow for *accidie*’. Shafer-Landau, *Moral Realism*, 179, (and 180 for emotional, neurotic and phobic susceptibilities impeding ‘rational’ agents); Schroeder, *Slaves of the Passions*, 167; Hurtig, ‘Internalism and Accidie’. Goldman avoids the objection of weakness of will by specifying that the relation is one between a reason and an element in the agent’s S. The result is a version of internalism similar to Weak Internalism: Goldman, ‘Reason Internalism’; Pettit and Smith also ‘stipulate away the possibility of irrationality’ (i.e., for them, ‘*akrasia* and the like’) and they note that Williams also allows for the possibility of *akrasia*: Pettit and Smith, ‘External Reasons’, 144–5.

⁴³ ‘On the second interpretation, there is no such condition, and the reason-sentence will not be falsified by the absence of an appropriate motive.’ Williams, ‘Internal and

WER: \neg [A has a reason to Φ only if: A_{BR+} could recognize the reason in question and upon doing so could Φ for that reason]

It will be convenient to abbreviate *the content* of the necessary condition in this formulation to ' $A_{BR+}RA$ '.⁴⁴

$A_{BR+}RA$: A_{BR+} could *recognize* the reason in question and upon doing so could Φ for that reason.

So abbreviated, WER states:

WER: \neg [A has a reason to Φ only if: $A_{BR+}RA$]

In contrast to WIR, WER clearly allows for the possibility that A has (or that there is for A) a reason to Φ where: \neg [it is a necessary condition for that reason that $A_{BR+}RA$].

Yet, as defined by Williams, WER is silent on whether it countenances the possibility of A having a reason to Φ where it *is* a necessary condition that $A_{BR+}RA$. Given this, it is open to interpret WER exhaustively so that it allows such reasons and it is more plausible for doing so because it does not then insist that there are *no* internal reasons.⁴⁵ Accordingly, WER maintains:

WER: There may be a reason for A to Φ where:

- 1) it is a necessary condition for that reason that $A_{BR+}RA$; or where:

External Reasons', 101[1]; 'The externalist view is that this is not a necessary condition, and that it can be true of A that he has a reason to ϕ even though A has no motivation in his motivational set that could, either directly or by some extension through sound deliberation, lead him to ϕ .' Williams, 'Internal Reasons and the Obscurity of Blame', 35[2].

⁴⁴ In this abbreviation, 'R' denotes 'recognize' and 'A' denotes 'act on'. The abbreviation may appear unorthodox, but we shall see that it is helpful.

⁴⁵ Williams writes: 'The whole point of external reason statements is that they can be true independently of the agent's motivations.' Williams, 'Internal and External Reasons', 107[1]. See also Ibid. 108[2].

- 2) \neg [it is a necessary condition for that reason that A_{BR+RA}].

While Williams does not define the terms ‘internal reason’ and ‘external reason’, and only uses either term occasionally, we know that he considers that external reason statements are either false or intended to mean something else and that there are only internal reasons. From this, we can infer that he intends the following definitions:

Internal Reason: a justifying reason that exists in virtue of corresponding (in some sense) to a true internal reason statement (i.e. a statement of the form: A has a reason to Φ only if A_{BR+RA}).

External Reason: a justifying reason that exists in virtue of corresponding (in some sense) to a true external reason statement (i.e. a statement of the form: A has a reason to Φ independently of whether it is a necessary condition that A_{BR+RA}).

The latter definition faces an interpretative issue about its scope and the meaning of ‘independently’. I need not justify my view on these issues in detail. Suffice it to say that the question of the definition’s scope is: whether for A having an *external* reason to Φ it is a necessary condition that: $\neg A_{BR+RA}$. In this regard, I see no reason to depart from the natural reading of External Reason which warrants the same exhaustive interpretation as that just given to WER:

External Reason*: a justifying reason for A to Φ where:

- 1) it is a necessary condition for this reason that A_{BR+RA} ; or where:
- 2) \neg [it is a necessary condition for this reason that A_{BR+RA}].

On this interpretation, *all* justifying reasons are external reasons. It is a further question whether a given justifying reason is *also* an internal reason. It will be if: it is a necessary condition for that reason that A_{BR+RA} . I continue to call justifying reasons that are both external and internal reasons in this way, *internal reasons*.

If, on this exhaustive interpretation, an external reason is *not* an internal reason, there are two further mutually exclusive classes that it may fall into:

1. internal reason
2. external-satisfying reason
3. pure external reason

An external reason will fall in the second class if it *satisfies* $A_{BR}+RA$ though (in contrast to internal reasons) it is not the case that: it is a *necessary* condition for the reason arising that $A_{BR}+RA$. Since it is *not* a necessary condition 'that $A_{BR}+RA$ ' for this external reason to exist, it is not an internal reason: it *satisfies* $A_{BR}+RA$, but it does not do so necessarily. I therefore distinguish such reasons as *external-satisfying reasons*.

An external reason falls into the third class on the exhaustive interpretation if it is known that it is neither an internal reason nor an external-satisfying reason. This means that it is known that it is not the case that ' $A_{BR}+RA$ is a necessary condition for the existence of the reason' and that the external reason does not, in any event, happen to satisfy $A_{BR}+RA$. It may be that the external reason 'does not happen to satisfy $A_{BR}+RA$ ' because it logically cannot or simply because it does not in the case at hand. Either way, I call such an external reason a *pure external reason*.⁴⁶

Williams contends that 'there are only internal reasons' and that 'there are no external reasons for action'⁴⁷ and it may be thought that these assertions favour a more moderate interpretation of External Reason. It may be

⁴⁶ There is a fourth class: silent external reasons. But I leave this class aside because it makes discussion unnecessarily complex. Silent external reasons consist of justifying reasons for which it is *unknown* whether it is true that ' $A_{BR}+RA$ is a necessary condition for the existence of the reason', and hence, it is unknown whether it is an internal reason, an external-satisfying reason or a pure external reason.

⁴⁷ Notably Williams qualifies his assertion that 'there are only internal reasons for action' by stating that this is a rather rough expression of his view. Williams, 'Internal Reasons and the Obscurity of Blame', 35[1], 41[1], 41[2].

thought, for instance, that if Williams adhered to External Reason*, then he *would* acknowledge that there are external reasons in the sense that there are external reasons that are also internal reasons (what I continue to simply call internal reasons). But, Williams is making these assertions as a way of expressing his view that WIR is true, and, if WIR were true, then it would be unnecessary and misleading to consider external reasons to exist, even in the sense associated with internal reasons. I therefore adhere to the interpretation in External Reason*.⁴⁸

On the question of the sense in which external reasons are ‘independent’ of A_{BR+} ’s motivations, Williams’ understanding, and the popular and perhaps unanimous understanding in the literature, is that WER is only true if in the *actual* world there are pure external reasons for A to Φ . Since more can be said in favour of WIR on this popular interpretation than other senses of independence, I adhere to it in what follows.⁴⁹ So understood, WIR avoids certain objections, but not all.

III The Conditional Fallacy

Recall that Williams considered what reasons A_{BR+} (instead of A) could recognize and respond to in order to avoid explaining why a given agent, A ,

⁴⁸ My interpretation here is equivalent with the version of externalism that Derek Parfit, E. J. Bond and Jonathan Dancy (following Parfit) maintain, and the interpretation Stephen Finlay argues for. That Parfit allows for a reason to be both an internal and external reason is evidenced in one of his replies to Williams’ argument: even though Jack has an external reason to do X this ‘does not imply that Jack has no internal reason to do X .’ Parfit, ‘Reasons and Motivation’, 100, 115; Bond, *Reason and Value*, chap. 2; Dancy, *Practical Reality*, 156; Finlay, ‘The Obscurity of Internal Reasons’, 17 col. 1. In the end, of course, clarity as to meaning is more important than terminology.

⁴⁹ Notice that this means that it is *not* enough to prove the truth of WER that an external-satisfying reason exists in the actual world because although by definition an external-satisfying reason ‘*could*’ exist without satisfying $A_{BR+}RA$, it does not mean that this reason could exist *in the actual world*—which is what we require.

does not always act as she has most reason to, due to lack of knowledge, weakness of will, apathy, intense passion etc. However, in this context (and in philosophy generally) conditional analyses are potentially vulnerable to a counterexample called ‘the conditional fallacy’.⁵⁰ This is the error of failing to recognize that in some cases where the antecedent of the *analysans* is true, the consequent of the *analysans* is false. This renders the conditional in the *analysans* false, thereby rendering the main conditional false as well. Thus, for example, Detective Lowe does not know where Augustine hid the revolver and in virtue of this she has a reason to search his property. However, if Lowe had full information and reasoned correctly (i.e. ‘Lowe_{BR+}’) she would already know that Augustine hid it underneath the abandoned church and hence, she would not have reason to search his property for it. Counterexamples with this form show that analyses of justifying reasons resembling analysis 1 below are in error:⁵¹

1. A has a reason to Φ in $C \rightarrow (A \text{ were cognitively enhanced with more knowledge } \{A+\} \rightarrow A+ \text{ would } \Phi \text{ in } C).$

They also disprove Williams’ necessary condition.⁵²

⁵⁰ Shope, ‘The Conditional Fallacy in Contemporary Philosophy’, 397.

⁵¹ Allan Gibbard, for example, puts this criticism to Brandt’s full information analysis of ‘rationality’. Gibbard, *Wise Choices, Apt Feelings*, 20–2. Analogously, in epistemology Alvin Goldman argues that internalist views face the problem that the average person may not be able to carry out large computational calculations. Goldman, ‘Internalism Exposed’, 219–20.

⁵² I must mention here that Williams himself charges John McDowell’s Aristotelian line of defence (that A has a reason to Φ only if the *phronimos* would have a reason to Φ) with the conditional fallacy. Williams, ‘Replies’, 190[2]–4[1]; cf. Williams, ‘Postscript: Some Further Notes on Internal and External Reasons’, 94[2]. Arguably this means that it would be uncharitable to interpret WIR in a way that exposed it to the same objection. However, Williams himself expressly denies that A must be able to be motivated to Φ for this reason. Instead he asserts that A having an internal reason ‘does not imply, as McDowell’s objection suggests, that the agent should be able to conduct the relevant deliberation in fact. Perhaps some unconscious obstacle, for instance, would have to be removed before he could arrive at the motivation to ϕ . Someone who claimed something of this sort would have to make good on the claim [...] This may not be easy to do, but the fact that there is an onus to be discharged here just marks the point, on the internalist’s view, that (R) [A has a

A. Railton & Smith

On this point, compare Peter Railton and Michael Smith's modified analyses:

1. **Railton's Internalism:** A has a justifying reason to Φ in $C \rightarrow A+$ would want A to want to Φ in C .⁵³
2. **Smith's Internalism:** A has a justifying reason to Φ in $C \leftrightarrow A+$ would advise A to Φ in C .⁵⁴

I use the term ' $A+$ ' here instead of A_{BR+} because Smith and Railton have different views of what idealization requires compared to Williams' description.⁵⁵

reason to ϕ] has a content which distinguishes it from other things that might be said about A .' Williams, 'Replies,' 188[2]. Other writers also interpret WIR as falling prey to the conditional fallacy: Robertson, 'Internalism, (Super)fragile Reasons, and the Conditional Fallacy'; Markovits, 'Internal Reasons and the Motivating Intuition', 150–3. Pettit and Smith, in contrast, state that Williams is silent on whether WIR can be interpreted on an 'exemplar' or 'advisor model' and since the advisor model avoids the conditional fallacy they maintain that Williams should be interpreted in this way. Pettit and Smith, 'External Reasons', 148. Also cf. Finlay, 'The Obscurity of Internal Reasons', 15 col. 1. Ultimately however, it does not matter whether WIR is interpreted in the way I suggest or as per Pettit and Smith. If WIR is a version of an advisory model then, while it avoids the conditional fallacy, it faces the same problem that Smith's theory faces: accounting for how the agent herself could always come to act on the reason in question.

⁵³ Railton, 'Moral Realism', 174. A biconditional does not apply here. Railton thinks that the reductive base for A 's justifying reasons are those facts about A and his circumstances that $A+$ would combine with his general knowledge in arriving at his views about what he would want to want were he to step into A 's shoes. He expressly denies that the reductive basis is the fact that $A+$ would have certain wants. 174, 175. Railton defines 'an individual's non-moral good' with a slight difference: 'what he would want himself to want, or to pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality.' Railton, 'Facts and Values', 54.

⁵⁴ Smith, *The Moral Problem*, 152; Smith, 'Internal Reasons', 110. Smith also presents a slightly different version of the analysis: If A has a reason to Φ in C , then $A+$ would want A to Φ in C . Smith, *The Moral Problem*, 181.

⁵⁵ Railton considers $A+$ to have: 'unqualified cognitive and imaginative powers, and full factual and nomological information about his physical and psychological constitution, capacities, circumstances, history and so on.' Smith considers that a fully rational agent has: all true non-moral beliefs, no false non-moral beliefs,

In each case, the consequent in the analysans accounts for the possibility that *A*'s circumstances may differ from *A*+'s circumstances. Each analysis thereby avoids the conditional fallacy. Thus, for example, whereas *Lowe*_{BR+} would have no reason to search Augustine's property, *Lowe*_{BR+} nevertheless appreciates that *Lowe* does not have such knowledge. Accordingly, on Railton's Internalism, *Lowe*_{BR+} would want *Lowe* to search Augustine's property and, on Smith's Internalism, *Lowe*_{BR+} would advise *Lowe* to search his property.

B. Johnson's Objection

Recently Robert N. Johnson argues that while the necessary relations that Railton and Smith identify succeed in avoiding the conditional fallacy, they only do so at the expense of one of the two virtues that makes Reason/Motivation Internalism (**P**₅) attractive.⁵⁶ The first virtue, according to Johnson, is that Reason/Motivation Internalism establishes a necessary relation between an agent's justifying reasons and the possibility that those reasons can be her explanatory reasons. The second virtue is that Reason/Motivation Internalism ensures that an agent's justifying reasons can be *her* justifying reasons from her personal point of view. Accordingly, Johnson provides a trilemma for proponents of Reason/Motivation Internalism: at least on every version so far put forward, the theory either:

1. commits the conditional fallacy;
2. (avoids the conditional fallacy but) fails to establish a necessary relation between an agent's justifying reasons and the possibility that those reasons can be her explanatory reasons.

'reasons correctly' and would also aim to achieve a systematic justification of her beliefs and desires. Railton, 'Moral Realism', 173–4; Smith, *The Moral Problem*, 156–61.

⁵⁶ Johnson, 'Internal Reasons and the Conditional Fallacy'; Johnson, 'Internal Reasons: Reply to Brady, van Roojen and Gert'.

3. (avoids the conditional fallacy but) fails to ensure that an agent's justifying reasons can be her justifying reasons from her personal point of view.

Thus, for example, while Johnson agrees that Smith's Internalism avoids the conditional fallacy, he argues that as it only requires 'that A^+ advise A to Φ in C ' the relation Smith posits will obtain when A Φ 's in C *accidentally* or for a reason other than the justifying reason in question. On this basis Johnson argues that Smith's Internalism does not ensure that A 's justifying reasons can be her justifying reasons from her personal point of view and Smith's Internalism is therefore impaled on the third horn of the trilemma.

Although Johnson does not argue for it, one may also be concerned that Smith's Internalism is also impaled on the trilemma's second horn. Smith's Internalism only entails 'that A^+ advise A to Φ in C ' and hence it does not entail either that **1**) it be physically or rationally possible for A to Φ in C or **2**) that A would, in any event, actually care for A^+ 's advice. Both reasons for skepticism are unpersuasive. In regard to the first reason I can only think that, given that A^+ has full information about the situation and reasons correctly, A^+ would know whether or not it is physically and rationally possible for A to Φ in C at t . In light of this, I do not see why A^+ would advise A to Φ in C at t if it were not rationally possible for A to do so.

In regard to the second reason, I find it difficult to believe that when an agent is reasoning faultlessly she would be indifferent to the advice of an epistemically advanced version of herself in the sense that she would not acknowledge that advice as a reason for her to Φ in C . Once one keeps in mind that, *ex hypothesi*, A^+ could not fail to have A 's interests in mind (as B or B^+ may fail to have A 's interests in mind) it is difficult to think of a reason that A could have for being indifferent toward such advice. The only reason I can think of is that A has received advice from a version of herself who is at an even greater epistemic and cognitive advantage than A^+ (perhaps A^{2+} has

knowledge of the future or a more coherent picture of the laws of physics, or is omnisciently rational). But in the absence of such advice or the knowledge that one will receive such advice, I consider that *A* would only be indifferent to *A*+'s advice if she were not reasoning faultlessly.

Richard Joyce rather deftly argues for this point in relation to Smith's Internalism, adding that just as it does not make sense to question practical rationality, it does not make sense for *A* to question what *A*+ would desire that she do, where *A*+ is only improved insofar as *A*+ has all relevant true beliefs, no false beliefs and reasons without error.⁵⁷ If *A*+ were 'improved' in other respects, such as being dispassionate as Firth's ideal observer is, then *A* may legitimately say 'so what if *A*+ desires I Φ '.⁵⁸ But that is not the case here.

C. Answering Johnson's Trilemma

Johnson's objection still threatens to impale Railton's Internalism and Smith's Internalism on the third horn of the trilemma. The problem with theories of Reason/ Motivation Internalism that fail to ensure that an agent's justifying reasons can be her justifying reasons from her personal point of view is that they entail that the theory is vulnerable to counterexamples of wayward causation. As Johnston discusses, according to Smith's Internalism, for example, there is a reason for *A* to Φ in *C* at *t* even if *A* can only Φ in *C* at *t* accidentally.⁵⁹

⁵⁷ Joyce, *The Myth of Morality*, 82–4.

⁵⁸ Ibid., 83. Railton also argues that *A* would care about what *A*+ would think about her own non-moral good. Railton, 'Moral Realism', 177–8. Cf. Rosati, 'Persons, Perspectives, and Full Information Accounts of the Good', 311–4; Sobel, 'Full Information Accounts of Well-Being', 793 n19.

⁵⁹ See Alfred Mele's work for discussion of three different kinds of wayward causation: Mele, 'Action', 338; Mele, *Springs of Action*, 207–10.

In light of this, Smith and Railton may either qualify their theories to avoid the third horn of Johnson's trilemma, or depart from Reason/ Motivation Internalism, thereby avoiding the trilemma altogether.⁶⁰ Each theory may be qualified in the same way:

3. **Railton's Internalism***: A has a justifying reason to Φ in $C \rightarrow A+$ would want A to want to Φ in C *for* that reason.
4. **Smith's Internalism***: A has a justifying reason to Φ in $C \leftrightarrow A+$ would advise A to Φ in C *for* that reason.

Recall that in this context the phrase 'for that reason' means that A Φ 's because she recognizes the consideration that gives rise to the justifying reason for her to act, as giving rise to that reason for her to act. We noted that WIR embraces this qualification (Cognitivism about Reasons) and it is justified in doing so. This is because if A has a reason to Φ in C *for* that reason, then A will not be responding to this reason *at all* if she Φ 's in C accidentally or for some other reason. This qualification therefore ensures that an agent's justifying reasons can be her justifying reasons from her personal point of view and that Smith and Railton's relations avoid the third horn of the trilemma.^{61/ 62}

⁶⁰ It is unclear from their recent published work whether either academic takes the latter view. It may appear from his 2006 paper with Philip Pettit that Michael Smith defends the existence of pure external reasons. While I think this is true, there is not strictly enough evidence to draw this conclusion. Pettit and Smith criticize WIR on the basis of Williams' belief that there is a close connection between blame and a reason to act. However, Reason/ Motivation Internalism, more generally, requires no such commitment. Pettit and Smith, 'External Reasons'.

⁶¹ J. David Velleman and Kieran Setiya also argue that we must understand agents 'acting' as 'acting for that reason' in this sense. However, Setiya denies that this must involve the agent taking her subjective justifying reason for acting as a 'good' or 'justified' reason for so acting (compared to an 'unjustified' subjective justifying reason). Velleman also adds this point in his 2009, (cf. his 1992). Velleman, 'The Guise of the Good', 100; Velleman, *How We Get Along*, 122–5; Setiya, *Reasons without Rationalism*, pt. 1 generally, p.60. Significantly, Mark Schroeder also observes that a psychological attitude of 'taking' something to be a reason does not automatically follow from the stipulative definition of this qualification of 'taking-something-to-be-a-reason'. However, he ultimately endorses it because this attitude draws the

Yet while Smith's Internalism, for example, avoids the second and third horns of the trilemma, it becomes unnecessarily complex in doing so. If $A+$ would only advise A to Φ in C for reasons that A could come to on her own, then there is no need to build the requirements of 'all relevant true beliefs' and 'no relevant false beliefs' into $A+$. Any reason to act that may arise from such enhanced knowledge is only filtered out again by the requirement that A be able to Φ in C for that reason.

The unnecessary complexity arises because if it is ensured that a rational agent's justifying reasons necessarily can become her explanatory reasons (the second horn of Johnson's trilemma) and her justifying reasons from her personal point of view (the third horn of Johnson's trilemma) then the conditional fallacy cannot arise in the first place. The conditional fallacy can only arise where one refers to some more idealized agent, $A+$, with qualities that A could not possess in the relevant circumstances (such as all true beliefs and no false beliefs). The fallacy then serves to show that this very difference between A and $A+$ could mean that $A+$ has justifying reasons that A does not.

right distinction between 'being caused to act' and 'acting for a reason' as well as providing an explanation for why it is right. Schroeder, *Slaves of the Passions*, 152–4. I think this is right: the difference between Donald Davidson's climber accidentally dropping his partner, and deliberately doing so (and hence, acting 'for the reason' that dropping his partner saves his own life), depends on the climber's conscious state of mind. I doubt that anything else could explain this.

⁶² Notably, Cognitivism about Reasons is not uncontroversial, but objections to it are sufficiently answered elsewhere. For example, Joseph Raz insists that mere 'conformity' to justifying reasons is sufficient. Yet his first argument for this view only briefly discusses an analogy between reasons to believe and act, and hence, requires further defence and his second and third arguments are actually consistent with Cognitivism about Reasons and Modest Internalism, as defined. Raz, *Practical Reason and Norms*, 178–82. Secondly, Cognitivism about Reasons is inconsistent with what Mark Schroeder calls the 'No Background Conditions view'. But one can consult his discussion on this matter: Schroeder, *Slaves of the Passions*, 27 n8, 28, 28 n10, 23–40 generally. Finally, Cognitivism about Reasons presupposes *Cognitivism about Intention*, the view, as I define it, that an intentional act, Φ -ing, involves a belief that one is Φ -ing. While this latter view is open to certain counterexamples (such as Donald Davidson's carbon-copier), I defer to Kieran Setiya's work on the issue: Setiya, 'Practical Knowledge', 389–92; Setiya, 'Practical Knowledge Revisited', 129–32.

But if it is stipulated that a rational agent's justifying reasons necessarily can become her reasons from her point of view and her explanatory reasons, then, with one exception, there is no difference between *A* and *A+* in which the conditional fallacy can arise to create trouble. The sole exception may be put forward as an objection in the following way:

A has a justifying reason to ensure that she reasons correctly or as correctly as she can, whereas *A+* does not have this justifying reason because, by definition, *A+* always reasons as correctly as *A* possibly could. So *A* may have a justifying reason to do something (namely, reason as correctly as she can) that *A+* does not have. So the purported necessary condition is falsified.

It is worth emphasizing that this is the only way the conditional fallacy may be put to a version of Reason/ Motivation Internalism that only distinguishes between *A* and *A+* on the basis that *A+* always reasons as correctly as *A* could in *C* at *t*. It also fails because there is no need to suppose that *A* and the construct, *A+*, know that necessarily *A+* always reasons correctly. Moreover, even if both parties knew that *A+* always reasons correctly and that *A* does not, this need not absolve *A+* of having a justifying reason to reason correctly where she does engage in such a process. Practical rationality provides all rational agents with reasons to reason correctly, whether they do so when required or not.

The result of avoiding the possibility of the conditional fallacy is that we do not need to consider 'what *A+* would *advise* or *want* *A* to do', we only need to consider 'what *A+* *could* do' where '*A+*' does not have all true beliefs and no false beliefs, but simply 'reasons as well as *A* could in *C* at *t*'. I understand 'reasoning as well as *A* could' here as involving the kinds of reasoning that Williams adds to standard means-end reasoning.⁶³ To some extent, reasoning 'well' or 'correctly' is therefore determined by the kind(s) of reasoning *A* is

⁶³ See this chapter, Section I.

employing. Thus, for example, if *A* is trying to identify the most efficacious means to a certain end, she will not be ‘reasoning as well as she could have’ if she does not do this, and she could have. The same is true if *A* is trying to resolve conflicts between elements of her *S_B* and she falsely believes she has done so when she could have actually done so.

However, ‘correct’ reasoning also adheres to other criteria. For example, it may involve reasoning that leads to a decision justified by subjective justifying reasons that are consistent with *A*’s beliefs. It may also mean that *A*⁺ concludes that she is insufficiently informed and that she should obtain more information before continuing. Finally, as discussed further in Chapter 5, an agent’s reasoning as well as she can may involve her taking account of information she gleans from certain events and external influences that she could not have recognized on her own. With this understood, *Modest Internalism* and *Immodest Externalism* consist in the following respective views:

Modest Internalism: There is a sufficient reason or conclusive reason, *R*, for *A* to Φ in *C* at *t* \rightarrow *A*⁺ could recognize *R* and *A*⁺ could be motivated to Φ in *C* at *t* for *R*.

Immodest Externalism: \neg Modest Internalism.

As we can see, Modest Internalism is a version of Weak Internalism and endorses Cognitivism about Reasons. In addition, the relevant sense in which it must be possible for *A*⁺ to recognize *R* and be motivated to Φ in *C* at *t* for *R* is cognitive and motivational possibility respectively. Furthermore, *Immodest Externalism* has, *mutatis mutandi*, the same exhaustive interpretation as WER and, similarly, the term ‘external reason’ under Immodest Externalism has, *mutatis mutandi*, the same exhaustive definition as that presented in External Reason*. Note, however, that Modest Internalism is limited to sufficient and conclusive reasons. This is because, by definition, insufficient pro tanto reasons provide insufficient justification to act on.

Hence, if after reasoning as well as *A* could in *C* at *t*, all that is left in favour of *A*+ Φ -ing is an insufficient pro tanto reason, *R*, to Φ , then *R* is an insufficient reason to Φ and it does not follow that, necessarily, after reasoning as well as *A* could, she could have some motivation to Φ for *R*. Perhaps in such circumstances *A* nevertheless has a very faint trace of motivation, but I leave this possibility aside here.

Finally, while there is disagreement over whether WIR consists in a constitutive, linguistic or conceptual claim about reasons, Modest Internalism makes a constitutive claim.⁶⁴ The right side of the conditional is a necessary condition for a consideration to be or give rise to a justifying reason to act.

The suggestion that Modest Internalism, so understood, might be true invites us to reconsider why Williams attributes all true beliefs and no false beliefs to *A*+. Recall that Williams' justification was twofold: 1. He explains that it is very odd to say that *A* has a reason to drink a glass of petrol when *A* believes it to be a glass of gin. It is more natural to say he does not have this reason, though he thinks he has. 2. Williams argues that internal reasons are not solely concerned with explanation. Rather, they are also concerned with the agent's 'rationality' and that 'any rational deliberative agent has in his *S* a general interest in being factually and rationally correctly informed' and 'to take the correct means to his ends'.⁶⁵

In regard to Williams' first justification we must firstly observe that in the scenario Williams sketches it must also be true that *A* rationally *cannot* come

⁶⁴ For discussion of this disagreement and defence of the view that WIR is a theory about the concept of 'a reason for action' see: Finlay, 'The Obscurity of Internal Reasons'. In a slightly similar vein, Derek Parfit argues that WIR is a theory about what people often mean when stating '*A* has a reason to Φ ' compared to a theory about the reason itself: Parfit, *On What Matters: Volume Two*, 269.

⁶⁵ See this chapter, nn14-16.

to know that the glass of gin is a glass of petrol. If *A* rationally *could* come to know this, then it may very well be said that *A* has no reason to drink the glass, though he thinks he has. In circumstances where *A* rationally *cannot* come to know that the glass of gin is a glass of petrol it may still be thought that it is more natural to say 'There is a reason for *A* not to drink the glass, though *A* does not know it.' I consider the linguistic evidence for external reasons in Chapter 5.

In regard to Williams' second justification it can firstly be said that Modest Internalism is also not merely concerned with *A*'s explanatory reasons. There is only reason for *A* to do what *A*⁺ could do, where *A*⁺ actually reasons as well as *A* could in *C* at *t*. Secondly, although it may be true that every rational agent has an interest in being correctly informed, as we shall see this in itself does not justify regarding *A*⁺ as having all true beliefs and no false beliefs. Chapter 5 considers this point in detail as we consider three arguments in favour of Modest Internalism. Discussion of these arguments also provides the basis from which we can identify where objections to WIR, as adapted for Modest Internalism, go wrong.

5

Modest Internalism I

In Chapter 4 we considered Williams' Internalism (WIR), one version of Reason/ Motivation Internalism (P_4 in The Relative Reasons Argument), and identified its commission of the conditional fallacy. We also observed that while Peter Railton's Internalism and Michael Smith's Internalism avoid this fallacy, they fall foul of Robert N. Johnson's trilemma. In contrast, Modest Internalism cannot commit a conditional fallacy and it is immune to counterexamples of wayward causation.

Modest Internalism: There is a sufficient reason or conclusive reason, R , for A to Φ in C at $t \rightarrow A^+$ could recognize R and A^+ could be motivated to Φ in C at t for R .

Here, necessarily, A^+ actually reasons as well as \underline{A} could in C at t . Immodest Externalism is defined as:

Immodest Externalism: \neg Modest Internalism.

This chapter presents three arguments in favour of Modest Internalism. In Section I, I contend that, contrary to Williams, in ordinary, non-philosophical conversation people rarely, if at all, make what Williams refers to as 'external reason statements'. As we shall see, these same considerations also apply to what would be classified as 'external reason statements' according to Modest Internalism. In Section II I argue that belief in pure external reasons is justifiably counterintuitive and in Section III I develop Williams' second

initial argument to show that there is not, in any event, a need to believe that there are pure external reasons. Chapter 6 turns to leading objections put to Reason/ Motivation Internalism, as adapted for Modest Internalism.

So structured, this chapter omits discussion of Williams' first argument for WIR—The Explanatory Argument.¹ This is because ultimately the argument assumes the truth of a premise that we must discuss anyway. Interpretation of Williams' arguments is controversial, but I consider it to have the following form:

The Explanatory Argument

1. 'If something can be a [justifying] reason for action, then it could be someone's [justifying] reason for acting on a particular occasion, and it would then figure in an explanation of that action.'^{2/ 3} ['The Explanatory Constraint']
2. Pure external reasons do not necessarily adhere to The Explanatory Constraint.⁴

¹ This argument originally arises at: Williams, 'Internal and External Reasons', 106[4]–107[3]. Williams then elaborates his discussion of this argument at 107[4]–110[1] and hence, these passages neither constitute nor include a separate argument.

² Ibid., 106[4]. For a similar account see: 102[4]. Williams makes a slightly different assertion in 1989: 'If it is true that A has a reason to ϕ , then it must be possible that he should ϕ for that reason; and if he does act for that reason, then that reason will be the explanation of his acting.' Williams, 'Internal Reasons and the Obscurity of Blame', 39[1]. His final account is at: Williams, 'Postscript: Some Further Notes on Internal and External Reasons', 93[1].

³ The modal phrase 'then it could be someone's [justifying] reason' in this constraint invites multiple interpretations. In particular, it *may* be interpreted as entailing that a justifying reason that is not 'presently someone's justifying reasons for acting on a particular occasion' as satisfying the constraint so long as that reason *could* be (in some sense) that person's justifying reason for acting on that occasion. Alternatively, this second modal term may simply be preserving the sense of modality as that expressed by the first modal term (and hence mean nothing) or it may be interpreted as distinguishing A_{BR+} 's justifying reason, compared to A 's justifying reason. Since I am not so concerned with interpretive issues I simply note that I regard the first interpretation as inconsistent with what Williams writes elsewhere and so I understand Williams according to either the second or third interpretation.

⁴ 'Pure external reason' was defined in respect to WIR, so I add here that in this argument the term 'pure external reason' is being applied *mutatis mutandi* to The

Argument for 2:

- a. If The Explanatory Constraint is true, then for pure external reasons to be justifying reasons at all, it must be that they could be someone's justifying reason for acting on a particular occasion.
- b. Yet, if this is true then they are, by definition, not pure external reasons.

C Theories claiming that there are pure external reasons, such as WER, are false.⁵

It is clear that Williams leaves the truth of **1** unjustified. I should point out however that, firstly, it is not clear whether he distinguishes what he calls the 'explanatory dimension'⁶ of justifying reasons from WIR.⁷ Secondly, whether or not this is so, by assuming the truth of **1**, it is not necessarily true that Williams is begging the question against WER. At least originally, Williams

Explanatory Constraint, as a different version of Reason/ Motivation Internalism. Nevertheless the argument is valid because The Explanatory Constraint is a more basic or more general than WIR, and so, WER denies that there are pure external reasons in this sense as well.

⁵ Similar arguments are given in favour of other internalist theses. Thus, in regard to internalism about mental content, Jerry Fodor argues that: **1)** mental states must have a content that can causally explain behaviour; and **2)** that broad content cannot causally explain behaviour and so, mental states must have narrow content. Fodor, *Psychosemantics*, chaps. 1–2. Similarly, in epistemology, John Pollock and Joseph Cruz argue that: **1)** epistemic norms must be directly accessible, otherwise epistemic agents cannot apply them; and **2)** externalist epistemic norms are not directly accessible, and the inability of epistemic agents to apply them therefore constitutes a conclusive refutation of externalist theories. Pollock and Cruz, *Contemporary Theories of Knowledge*, 133–4.

⁶ Williams, 'Internal and External Reasons', 102[4], 106[4].

⁷ It is plausible that Williams means that justifying reasons have some kind of explanatory dimension such that, for a justifying reason to exist, it must be possible for it to be an explanatory reason for someone's acting on it (or for the agent in question acting on it). Williams has his own way of fleshing out the details of this explanatory dimension (*A*'s idealization to *A_{BR+}*) and he provides arguments for these qualifications, the result of which is WIR. But the explanatory dimension he refers to differs from WIR. For this reason, David Sobel and Russ Shafer-Landau err in their objection to WIR that this more general explanatory constraint on justifying reasons just is WIR. Sobel, 'Explanation, Internalism, and Reasons for Action', 223; Shafer-Landau, *Moral Realism*, 179.

takes this premise to be common ground among philosophers.⁸ In any event, for our purposes the significant point is that this premise now requires defence and so, even if The Explanatory Argument were adapted to defend Modest Internalism, we would still have to prove the truth of Modest Internalism as premise 1, which is what we turn to now anyway.

I External Reason Statements

Recall that while Williams avers that people make external reason statements, he denies that there are external reasons. These statements are justifying reason statements made without the intention that the truth of the statement is dependent on its compliance with a certain necessary condition. The use of such statements is not determinative of the question of whether

⁸ This is evident in his original paper. After stating The Explanatory Constraint, Williams identifies that an external reason statement could not, without more detail, explain *A*'s Φ -ing for that reason because the statement is by definition true independently of *A*_{BR}'s motivations. Accordingly, Williams considers how an external reason *could* comply with The Explanatory Constraint and observes that 'nothing can explain an agent's (intentional) actions except something that motivates him so to act.' Hence, at this point Williams is trying to identify how WER could be true when writers already accept The Explanatory Constraint. Williams acknowledges that 'there are various means by which the agent could come to have the motivation and also to believe the reason statement'. She may, for example, be persuaded by rhetoric. But, Williams asserts, these are 'the wrong kind of means to interest the external reason theorist.' What 'the external reason theorist essentially wants', Williams claims, is 'that the agent should acquire the motivation *because* he comes to believe the reason statement, and that he should do the latter, moreover, because, in some way, he is considering the matter aright.' Ultimately then, Williams considers that proponents of WER not only want to comply with The Explanatory Constraint, they want to comply with it *in a certain way*. Williams, 'Internal and External Reasons', 108[4]–109[1] (emphasis present).

Significantly, Stephen Finlay explains that Williams may have assumed the truth of The Explanatory Constraint because explanatory reasons were not commonly distinguished from justifying reasons at the time, particularly in light of Donald Davidson's and, I may add, Gilbert Harman's work in the 1960s-1970s: Finlay, 'The Obscurity of Internal Reasons', 13 col. 2–14 col. 1.

Indeed, more recently, Jonathan Dancy and Richard Joyce also assert and leave undefended the claim that there is a general explanatory constraint on justifying reasons. Dancy, *Practical Reality*, 101; Joyce, *The Myth of Morality*, chap. 5.

there are external reasons, although they are considered to provide linguistic evidence in favour of the existence of reasons of this kind.⁹ Similarly, lack of linguistic evidence for external reasons provides support for, but does not prove, the view that there are no external reasons. In fact, I consider linguistic evidence to bear less probative value than it is often taken to possess because it arises in pre-reflective contexts. Nevertheless, since there is either an assumption or consensus in the literature that people make external reason statements, and, to a lesser extent, since other academics consider that linguistic evidence is more valuable than I do, I discuss this matter in more detail, noting three considerations that cast doubt on the existence of such linguistic evidence. These considerations do not prove that people do not make pure external reason statements because proof or disproof of such is quite elusive. It would require knowing what a significant percentage of the population mean when they make assertions about justifying reasons which may be indeterminable. Nevertheless, taken together, these considerations provide significant evidence for doubting that we possess the kind of linguistic evidence for people making pure external reason statements that Williams and other writers assume we have. The result not only casts doubt on the linguistic evidence favouring Immodest Externalism, but also entails that we have good reason to believe that the *only* clear linguistic evidence we have favours internal reasons.

In Chapter 4 we saw that, according to Williams, statements of the form 'A has a reason to Φ ' and 'There is a reason for A to Φ ' must meet a necessary condition to be true ($A_{BR}+RA$). Despite this, Williams believes that these types of statements are asserted (and interpreted) in two senses:

⁹ However, compare such statements provision of evidence for the *actuality* of what their content represents, instead of its *existence*: Skorupski, *The Domain of Reasons*, 425–6.

1. a sense that is intended to comply with this necessary condition (an internal reason statement); and
2. a sense that is not intended to comply with this necessary condition (an external reason statement).¹⁰

To this we can add that, where it is the intention of the speaker of the reason statement that 'the statement is true even though *A* could not rationally Φ for the reason suggested', the reason statement is a *pure* external reason statement. It is a statement asserting that there is a pure external reason. Since Immodest Externalism maintains that there are one or more pure external reasons in the actual world, I focus on statements that are pure external reason statements according to Immodest Externalism, not WER.

It is difficult to identify such statements. For a reason statement to be a pure external reason statement it must have been uttered or written while the speaker or writer *does not believe* that one or more of the rational agents in question can in some sense be motivated to Φ with that reason in mind. (The phrase 'in some sense' captures the sense in which there is a necessary relation between justifying reasons and motivation according to Modest Internalism). Significantly then, whether a statement is a pure external reason statement is not evident from the mere fact that a reason statement was made. Nor is it enough (as I think it is usually assumed to be) that a reason statement is made in circumstances in which it is true that one or

¹⁰ While Williams defines the terms 'reason statement', 'internal interpretation', and 'external interpretation', he does not define the terms 'internal reason statement' and 'external reason statement'. It may then be that the denotation of 'internal reason statement', for example, is confined to: 'a reason statement that expressly shows that the internal interpretation is intended from the wording of the statement itself'. Alternatively, it may be that the denotation of 'internal reason statement' is: 'a reason statement on the previous interpretation or a reason statement, the meaning of which is to be understood according to the internal interpretation, though the statement's wording provides no express indication of this'. I follow the latter, broader, interpretation in this thesis as Williams expressly provides that: 'People do say things that ask to be taken in the external interpretation.' Williams, 'Internal and External Reasons', 106[2].

more of the rational agent(s) the subject of the statement, could not rationally have Φ -ed for the justifying reason in question. In accordance with Williams' definition, and one that we should follow in this regard, the speaker herself must make the reason statement *while* believing that one or more of the rational agent(s) in question could not recognize the justifying reason or otherwise be motivated to act in light of it.

Consequently, it is not enough for a philosophical distinction to maintain that it *seems* like there is a relevant difference between reason statements of the kind:

a) 'There is a reason for A to Φ '; 'There is a reason for you to Φ '; 'There is a reason for everyone to Φ ' (and perhaps 'Everyone has a reason to Φ ')
 and reason statements of the kind:

b) ' A has a reason to Φ '; 'You've got a reason to Φ '

such that statements of kind **a)** imply that the reason exists independently of whether A believes it to exist or independently of whether she could act for that reason.¹¹ The speaker or writer's intended meaning is more important.

Yet, that reason statements are uttered while the speaker or writer believes that the agent could not rationally be motivated to Φ for the relevant reason is a more difficult proposition to sell. Part of the difficulty is that people may have one of several, different intended meanings in mind when making such statements:

1. That a certain agent, A , has a reason to Φ and A can Φ .
2. That a certain agent, A , has a reason to Φ whether or not A can Φ

¹¹ Stephen Finlay also makes this point: Finlay, 'The Reasons That Matter', 15.

3. That a certain agent, *A*, has a reason to Φ even though *A* cannot Φ .
4. (Nothing more is consciously intended).¹²

Consider then the following ordinary reason statements:

- a. Everyone has a reason why they participate in the race, whether they run to live or they live to run. They run with mothers and fathers, with brothers and sisters, with husbands and wives. They come to outrun their demons and their diagnoses.¹³
- b. The movie industry has a reason to celebrate at this year's Academy Awards thanks to Avatar's record box office billions.¹⁴
- c. My husband has given me his word, sir,' said Madame de Villefort; 'you have just seen him resolve to keep it when he has everything to lose, and surely there is more reason for his doing so where he has everything to gain.¹⁵
- d. "It's easy to write off California - I know I have - but I think now more than ever there's a reason to pay attention," says Juliette Pope, wine director at Gramercy Tavern, one of the city's top wine destinations.
What's caught her imagination is a growing subset of winemakers "dialing back" on ripeness and making more restrained wines.¹⁶
- e. What is the best program to get the best quality sound out of a WAV file? I have Windows Vista. [...]
For just playing WAV files, use whatever software has an interface you're comfortable with - Windows Media Player, iTunes, etc. - I don't think there's a reason to pick one over the other when it comes to audio quality.¹⁷
- f. We all have a reason to celebrate! Lady Gaga's album "Born This Way" is released on May 23rd and since we all love Gaga's music and what she stands for we decided our next contest should be inspired by her!¹⁸
- g. Whether its sport, music or food, everyone has a reason to visit Knysna this winter.¹⁹

¹² I take the words 'can' and 'cannot' here as having their ordinary, non-philosophical, ambiguous meanings of: 'ability', 'capacity' and 'inability', 'incapacity' respectively.

¹³ Robbins, 'Supporting Cast', 101.

¹⁴ Schembri, 'The Academy Awards Rebooted'.

¹⁵ Dumas, *The Count of Monte Cristo*, 490.

¹⁶ Baiocchi, 'N.Y. Takes a New Look at California Wines'.

¹⁷ walternewton, 'Best Program to Play WAV Files In?'.

¹⁸ BreakoutBand, 'We All Have a Reason to Celebrate! Lady Gaga's...'

- h. No matter what your current situation is, you have a reason to be grateful! In fact, if you dig deep enough you can find enough reasons to be grateful for to change your life in a instance [sic]. Just the fact that your [sic] alive and breathing is a reason to be thankful!²⁰
- i. Everyone has a reason to quit smoking. What's yours? [Title]
[...] Nobody said quitting is easy but it helps when you know why you want to do it. Here are 34 real life examples of individuals who successfully quit smoking. If they can do it, so can you.²¹

Statements **a-c** above can quite straightforwardly be dismissed as not being pure external reason statements. From the context of statement **a**, the runners already know of their justifying reason for running. It is also reasonable to consider that in statement **b** the movie industry knows of their success and that in statement **c**, Madame de Villefort's husband knows of his reason to keep his word, because, as in statement **a**, he has already done so before. Moreover, the implication in the passage in **c** is that he would be aware of the fortunate consequences of doing so where he has everything to gain. Statement **d** is clearly addressed to wine connoisseurs and while it may have been intended as asserting that the reason applies even to those connoisseurs who could not recognize it, this appears to be unlikely. It is unlikely that Juliette Pope has access to facts about whether a particular connoisseur could learn of the reason to pay attention to Californian wine makers and it is unlikely that she has an interest in identifying such facts about other connoisseurs herself or making a claim about what such connoisseurs have reason to do irrespective of what they can come to learn. Rather, it is more natural to understand her as merely playing a role of informing connoisseurs about the reason to pay attention to California. This is not to say that all such connoisseurs could have recognized this reason. Rather, it is just to say that, in light of this, the statement is unlikely to have been expressed with the

¹⁹ Knysna Tourism, 'Knysna Oyster Festival'.

²⁰<http://www.successdemandsaction.net/2010/08/quickest-way-to-change-your-life.html>

²¹ Health Promotion Board, Singapore Government, 'Everyone Has a Reason to Quit Smoking. What's Yours'.

intent that certain connoisseurs have this reason even if they cannot recognize and act on it. Statement **e** is an illustration of a negative reason statement: it states that there is no such reason.

Statements **f-i** purport to apply to everyone and so they are, *prima facie*, better candidates for pure external reason statements. However, there is difficulty interpreting statements **f** and **g** as being made with the *genuine* intention that they *do* apply to everyone. More plausibly, the writers here are not thinking so much about whether this is the case at all, for they do not really care. Rather, in **f** the writer's concern is instead focused on emphasizing that Lady Gaga's music is great and that *Born this Way* will soon be released. The writer's reference to 'We all' in the sentence 'We all have a reason to celebrate!' is either a way of emphasizing how great her music is and/ or an abbreviated assertion about what *everyone who loves her music* has a reason to do. A similar point applies to **g**. As such, both statements are apposite illustrations of bluffing. Furthermore, it is unclear that the writer of statement **h** intends it to apply to someone who cannot see any reason to be grateful for anything—even for being alive or to someone who cannot feel grateful. It is entirely possible that the writer thinks that everyone can feel grateful and can recognize a reason to be grateful, as indicated by the suggestion that you can change your life in an instant. Finally, statement **i** appears to be addressed to people who already want to quit smoking, and hence, people who already recognize the relevant justifying reason (whatever this may be) for doing so.

Of course, statements **a-i** are not representative of all types of reason statements that are made but the first point of these illustrations is to merely ground discussion in live examples. The second point is that it is difficult to identify pure external reason statements because it requires interpreting a speaker or writer's intentions. Indeed, we have already seen the difficulty of this in Chapter 3 in regard to agents' evaluative attitudes.

Williams' own argument for there being external reason statements is as follows:

People do say things that ask to be taken in the external interpretation. In James' story of Owen Wingrave [...] Owen's father urges on him the necessity and importance of his joining the army, since all his male ancestors were soldiers [...] Owen Wingrave has no motivation to join the army at all, and all his desires lead in another direction: he hates everything about military life and what it means. His father might have expressed himself by saying that *there was a reason for Owen to join the army*. Knowing that there was nothing in Owen's S which would lead, through deliberative reasoning, to his doing this would not make him withdraw the claim or admit that he made it under a misapprehension. He means it in an external sense.²²

Significantly, Williams states that Owen's father does not actually make such a reason statement, no doubt because he is dead at the commencement of the story.²³ Certain other characters try to convince or talk about trying to convince Owen to join the army, but they do not make statements about reasons of any kind. Hence, Williams relies on the point that it might be imagined that the characters said this.²⁴ But we have to be careful. A philosopher contending that we do make external reason statements cannot conjure up such statements to prove that they are made without begging the question. She can only maintain that such statements already exist and that philosophers are *only* using their imagination to measure the plausibility of believing that they already exist. Three considerations tell against the likelihood of this.

A. Consideration 1: Implausibility of Content

It is doubtful that people give voice to reason statements where they know with certainty or a high level of confidence that it is not rationally possible

²² Williams, 'Internal and External Reasons', 106[2] (emphasis present). See also: 111[2]. His clearest statement of this view appears at: Williams, 'Postscript: Some Further Notes on Internal and External Reasons', 93[2].

²³ Williams, 'Internal and External Reasons', 110[5]–111[1].

²⁴ Id.

for A to Φ in C for the reason in question because, in light of *knowing* of this *very inability* of the agent, it is counter-intuitive to consider that there is this reason for the agent or that the agent has this reason. It would require believing that in some cases when people assert: “There is a reason for A to Φ ” the intended meaning of such a statement is one or both of the following (where the first entails the second but not *vice versa*):

1. There is a reason for A to Φ in C at t even though A will never know this.
2. There is a reason for A to Φ in C at t even though A will not be motivated to Φ in C at t for that reason.

It is difficult to accept that people would make assertions with such an intended meaning either consciously or on some more subtle level. It has the implication that, while it may not defy the laws of physics for A to Φ in C at t , it is physically impossible in a deterministic sense for A to Φ in C at t for that reason. This would mean that the intended meaning of the reason statement is equivalent to: ‘There is a reason for A to Φ in C at t with this reason in mind even though it is predetermined that she will not Φ in C at t with this reason in mind’.

Where the content of someone’s belief is equivalent to, but slightly varies from, that stated above, we may have a different answer. For example, consider the following content of someone’s belief:

1. There is a reason for A to Φ in C at t even though it is not possible for A to *know* that she has this reason; or
2. There is a reason for A to Φ in C at t even though it is not possible for A to *be motivated* to act on it.

In particular:

3. There is a reason for *A* not to kill her target at *t* because assassination for financial gain is immoral, even though it is not possible for *A* to recognize this reason at *t* or act on it at *t*.

Is it plausible to imagine a rational agent making a genuine reason statement while knowing this? One's intuitive answer here might be: Yes. But an affirmative answer may be directed by a concern that one could be mistaken about what *A* could have done in conjunction with wanting to ensure that *A* is held blameworthy. Evidence of this derives from the fact that when the statement is phrased such that *A will not* recognize the reason, as in statements 1 and 2, the merit of blaming *A* is to some extent lessened in virtue of this phrase emphasising that, of the course of possible events that may take place from here on, none consist in *A* seeing and acting for the reason in question, irrespective of what *A* wants to do or tries to do. In virtue of this we also get a different intuitive answer to the question of whether the following statement is true:

- 3'. There is a reason for *A* not to kill her target because assassination for financial gain is immoral, even though *A will not* recognize this reason or act on it.

To make the point explicit, suppose the following is true:

- It is not possible for amoral assassin *A* to see and act on reason *R* (a reason *not to kill her target* in light of the consideration that most people think that the act is immoral/ that it is immoral).
- *A* may or may not be caught.
- If *A* is caught, she will be held responsible.

While aware of these facts, we can ask whether a speaker would genuinely assert and believe: "There is a reason for *A* not to kill her target because it is immoral"? or "*A* has a reason not to kill her target because it is immoral"? It is entirely possible, though a little more difficult than usual, to find an intuitive answer here because we, as philosophers, must imagine ourselves

in the shoes of a layperson who, to answer this question, we think would and must imagine herself in the shoes of the assassin. Yet firstly, I believe that we do have the intuition that there is *not* a reason for *A* here. (My own intuitive response is: 'No, she's not going to see this reason.')

Secondly, the very fact that we think that a layperson would put herself in the shoes of the assassin shows that whether they would make the reason statement in this case is dependent on what the assassin would think is rational to do and what the assassin would do. Hence, if it is clearly stipulated that *A* will *not* see and act on the reason in question, there is less impetus for blaming *A* and correspondingly more doubt that people would make external reason statements about *A* in this situation. Of course, *A* may still be blamed and *A* may still be blameworthy, because the dispositions of thought and action that *A* has cultivated up to this point may be *A*'s responsibility. But the question of whether there is a reason for *A* not to kill her target in *C* at *t* is seemingly no longer determined *entirely* by a concern for the agent avoiding responsibility and as a result it is doubtful that one would assert that *A* does have this reason.

In light of this, it is far less plausible to suppose that people utter or write reason statements while genuinely believing that *A* could not see the reason to act in question (or if she could see it, could not be motivated to act for that reason). This does not suffice to prove that external reason statements are not made, but it does support the weaker conclusion that they are relatively rare. The rarity of such statements being uttered is significant because if the utterance of an external reason statement is rare then it is questionable that there is *sufficient* linguistic evidence to infer that we are committed to there being external reasons.

B. Consideration 2: Rarity of Knowledge

The second consideration evidencing the absence of pure external reason statements arises from epistemic concerns and supports the claim, just mentioned, that pure external reason statements are rare, if present at all. One may doubt that a speaker could ever or often *know* with certainty or a high level of confidence that either an amoral hired assassin or indeed a morally perverse agent *cannot* come to see certain considerations as reasons.²⁵ For example, perhaps it could be doubted that a morally perverse agent could *not* see a reason or come to see a reason to be genuinely merciful to her victim. Perhaps, even though a morally perverse agent and an amoral assassin have rather uncommon psychological dispositions, they are still rational agents in the sense that they respond to reasons that they believe exist. Furthermore, we might think that, insofar as they are human, they could come to see a reason to be merciful to a victim or target. Perhaps it is hard to be certain or highly confident that they could *not* do so.

While there are legitimate grounds for raising this doubt, it is not convincing enough. It *might be* that a person could come to know that a certain hired assassin could not see a reason not to kill her target because it is immoral. However, firstly, even if a speaker did know with a high level of confidence or certainty that an amoral agent is unable to see certain considerations as reasons, such knowledge must be rather rare. It must be rare because, as we have discussed, we cannot observe, intuit or otherwise directly acquaint ourselves with the mental and emotional states of others at work and determine what the agent can accomplish on this basis. The best that we can do is infer the existence and status of such states, and yet, we do not always have the information to make such an inference as often we are only familiar with people from a distance or, if we know them well, we can form such a

²⁵ Richard Joyce argues that we can conceive of morally perverse agents and that indeed, there are historical cases of such individuals. Joyce, *The Myth of Morality*, 19.

belief on the basis of probative (but not conclusive) evidence. No doubt, as a result, people's responses sometimes surprise us.

In addition, the very fact that it is not rationally possible for A to Φ for a certain reason may in itself mean that A has such a nature and set of dispositions that we cannot rely on judgments of generality and that we have difficulty imagining ourselves in A 's position at all. If, for example, alien beings with rational capacities (that is, capacities to reason in accordance with certain principles) arrive on Earth demonstrating a lack of concern for human welfare, I doubt that we would say that 'there is a reason for them not to annihilate humanity'.²⁶ Rather, I think that we would realize that we are in no position to say what they have reason to do or not do because we have no idea what/ who they are at all.²⁷

This may also be true of the kind of amoral and perverse agents which externalists may cite as cases in which we may know with certainty that the agent could not see or could not act on the reason in question. It is often difficult enough to firmly understand the world from another person's point of view. Attempting to comprehend and perceive the world from the point of view of a perverse moral agent may require such a radical change in one's perception and understanding of the world that one may always be uncertain about whether one has completely and accurately grasped that point of view or not.

Given then the implausibility of people uttering external reason statements in light of their content, and the rarity with which people possess certainty or a high level of confidence that A could not see the reason to Φ for certain

²⁶ The illustration is from: Harman, 'Moral Relativism Defended', 5.

²⁷ This is also a point R. M. Hare makes and illustrates with a case of trying to imagine oneself in the position of someone who is mentally ill. Hare, *Freedom and Reason*, 127.

considerations, the cases where people utter external reason statements, if any, are unlikely to arise sufficiently often to constitute the linguistic evidence externalism relies on. This means that the probative value of linguistic evidence externalists cite is weakened in proportion to their rarity. Indeed, there must be a tipping point at which the rarity of such utterances in fact starts to favour the *non-existence* of pure external reasons. Without a tipping point, consistency would require us to believe that linguistic evidence favours the existence of Pricilla pink gremlins were someone to genuinely report sighting them and nothing to the contrary is uttered. It might be argued that if someone did genuinely see Pricilla pink gremlins this would afford *some*, but not necessarily sufficient, linguistic evidence for their existence. But the point is not that genuine utterances cannot provide some evidence for our commitment to something. They may or may not do so. The point is that, overall, the rarity of such utterances means that there is insufficient linguistic evidence to suppose that we are committed to the existence of justifying reasons of this kind.

C. Consideration 3: Other Expressions

Thirdly, in cases where the speaker knows with certainty that it is not rationally possible for A to Φ , while we can easily imagine the speaker asserting ' A should Φ ' or 'It would be better if she did' we are less likely to imagine the speaker claiming either that: 'There is a reason for A to Φ ' or ' A has a reason to Φ '. This may be because the former assertions can easily be read as solely expressing the force of a normative standard whereas it is less clear that this is so where it is asserted that someone has a reason to Φ . It sounds as though the truth of the assertion 'There is a reason for A to Φ ' is more personally bound up with A than the statement ' A should Φ ' which may instead be read as expressing the evaluator's personally accepted normative standard.

If this is right, then the lack of linguistic evidence for pure external reasons undermines the argument for existence of those reasons. Moreover, this shows that we do not need to invoke pure external reasons to explain our use of language and I consider an argument from parsimony in Section III below. Of course, lack of linguistic evidence for pure external reasons does not refute the possibility of their existence. In 1801, we did not talk about antibiotics, genes and magnetic fields, but they existed. Similarly, a reverse error theory could be true of pure external reasons: our language does not presuppose or implicitly entail the existence of something significant that *does* exist.²⁸ But this is not true. Lack of linguistic evidence only slightly weakens the argument for the existence of such reasons, but there is more valuable evidence against them.

II *Intuitive Evidence*

An explanatory constraint of some kind for the truth of the statement 'There is a reason for A to Φ ' has intuitive appeal: there is no point in claiming that this statement is true if there is not some sense in which A could Φ . Specifically, however, for Modest Internalism there needs to be intuitive support for believing that:

1. There *cannot* be a conclusive reason for A to Φ if:
 - i) A could not see the reason to Φ (and therefore could not be motivated to Φ for that reason); or if
 - ii) A could see the reason to Φ but could not be motivated to Φ for that reason.

²⁸ A different kind of reverse error theory arises where one does not talk seriously about something that one knows that *others* believe to exist, but one does not, oneself, believe it to exist. For example, theists could claim that a reverse error theory is true for atheists: they do not believe that God exists when they talk about God, and yet, God does exist.

On the other hand, Immodest Externalism requires support for believing that:

2. There is a conclusive reason for A to Φ even though:
 - i) $A+$ could not see the reason to Φ (and therefore could not be motivated to Φ for that reason); or
 - ii) $A+$ could see the reason to Φ but could not be motivated to Φ for that reason.

When it is kept in mind that $A+$ reasons as well as A could in the relevant circumstances and at the relevant time, I consider that there is greater intuitive support for Modest Internalism because I cannot help but ask, 'What more can $A+$ do?' or 'What more could anyone do?' Conversely, if it were not possible for $A+$ to see the reason to Φ and to Φ for that reason, I find myself wondering what the point of it (the reason) is. Implicit in my own intuitive response to each statement is some version of The Explanatory Constraint,²⁹ but philosophers' intuitions vary and I am suspicious of the credibility of intuitions in general. If there is intuitive evidence favouring Modest Internalism, it must have stronger foundations.

²⁹ It is certainly not true, as Mane Hajdin claims, that as there appear to be external reasons, proponents of Reason/ Motivation Internalism have the burden of disproving their existence. Firstly, once it is clarified that this is a reason A could not recognize and act on, the intuition is tenuous. Secondly, even if the intuition is present, it is to be weighed against the intuition favouring The Explanatory Constraint noted above. Hajdin, 'External Reasons and the Foundations of Morality: Mother Teresa versus Thrasymachus', 435. Similarly, it is not true, as Richard Paul Hamilton suggests, that internalists have the burden of proof because we often talk of a person having legal reasons to Φ in an externalist sense. Williams is not solely concerned with moral reasons: he is concerned with reasons to act generally and reasons to act in light of legal sanctions are not sufficiently special to absolve the application of the two points noted in reply to Hajdin. Hamilton, 'Might There Be Legal Reasons?', 435.

A. Limitations on Reason Ascriptions

As a starting point, we can notice that there are at least *some* intuitive limits to our willingness to consider that it can be true that 'There is a reason for *A* to Φ '. In coming to the conclusion that rational possibility is the most viable sense of 'possibility' in Reason/ Motivation Internalism, we already noted Russ Shafer-Landau's argument that 'physical possibility' as the relevant sense of possibility is too weak and too strong.³⁰ It is too strong, he argues, if determinism is true, for then we would not want to say that '*A* has a reason to Φ ' when Φ -ing involved doing something that *A* was not going to do anyway. For the same reason, we also do not intuitively consider that: '*A* has a reason to create a fire by imagining one', even where it is true that doing so, if he could, would save someone's life.³¹ The point is not limited to morally significant cases. Intuitively, we do not think that '*A* has a reason to take a photograph of the Palace of Nations (Switzerland) right now' when *A* is presently in Vientiane, Laos. So far as we know, the time frame given makes this physically impossible.³²

In the same vein, suppose that a demon credibly threatens to hurt *A* at *t* unless she: **a)** squares the triangle; or **b)** cones the square and does not cone the square; or **c)** cones the square and does not cone the square while creating fire from her imagination in Vientiane and taking pictures of the

³⁰ See Chapter 4, I.

³¹ I leave aside any exceptions to the physical impossibility of such a feat arising from the aid of technology. If there are any such exceptions we may assume that they do not apply in the case in consideration.

³² John Skorupski gives a variation of the evil demon case in buck-passing accounts and outlines the positions of both views on the issue of whether '*A* having a reason to Φ ' implies '*A* can Φ '. But he remains neutral on this point. The argument for the opposing view, Skorupski notes, is that 'ought implies can' only applies to moral cases. I can have reason to believe something I cannot believe, so why not reason to do what I cannot do? Thus if a billionaire offered me a million dollars to square the circle I would have a reason to do it. Skorupski, *The Domain of Reasons*, 92. The latter argument depends on an analogy between theoretical and practical reasons which I briefly discuss in Chapter 6, IV.

Palace of Nations, at t^2 or **d**) does not have reasons to believe, feel, and act. Again, intuitively I doubt that such reasons exist.

However, in general I deny that a principle derived from a given intuitive response deserves the degree of authority it is often accredited. Firstly, the intuitive response may be wrong. Secondly, one may derive the wrong principle from an intuitive response. Since intuitive responses are pre-reflective, a principle derived from one may be inconsistent with intuitive responses one has or would have toward either the same scenario described in richer detail, different scenarios, or, more generally, developed theory.³³ Thirdly, even where the scope and content of the principle derived is correct and the intuitive response is correct, it is still not clear that such principles should be regarded as having the degree of authority they are granted.

What then should we say about reasons for physically and logically impossible acts? The principle to derive from the intuitive responses above is:

where an agent knows, with certainty, or as much certainty as she can have, that an act, or the state of affairs an act is intended to bring about, is logically or physically impossible in the actual world, then there is no justifying reason for the agent to perform that act.

One objection that might and perhaps should be put to inferring that this is the appropriate principle to derive from intuitive responses in the above cases is: it is not that there is *no* reason to act; it is just that it is a very weak reason. Mark Schroeder argues that this is the case on the basis of illustrations wherein the relevant act is physically possible, as we see when other considerations are, counterfactually, introduced.³⁴ Yet, where the act

³³ For one example, see Felipe De Brigard's discussion of Nozick's well known thought experiment: De Brigard, 'If You Like It, Does It Matter If It's Real?'

³⁴ Schroeder, *Slaves of the Passions*, 92–7.

itself or the state of affairs intended to be brought about is, at least by stipulation, logically impossible or a violation of the laws of physics, so long as we are still discussing the actual world, no further considerations when counterfactually introduced *could* logically show that that act, or intended state of affairs could now occur. If they could, then it is only by our misunderstanding what 'logically impossible' means or what the laws of physics are, in which case, the error is one of stipulation.

There is also independent support for this derivative principle. By definition, a justifying reason for *A* to perform an act, Φ , in *C* at *t* is a consideration that favours Φ -ing in *C* at *t*. 'Favours' denotes a normative relation between the consideration and Φ -ing. But the state of affairs 'squaring the triangle' and the act that brings this about cannot logically exist in the actual world, and so, there is no such act or state of affairs *to be favoured* in the actual world (except under the guise of false belief). In turn, the favouring relation cannot exist as a relation that favours something in the actual world (absent false belief). One can knowingly refer to non-existent possibilities, but not 'non-existent possibilities *in the actual world*'.³⁵ Furthermore, without a favouring relation there cannot be a justifying reason and this same rationale applies for acts and states of affairs that defy the laws of physics in the actual world.

The question then is whether such limitations extend to cases where it is rationally impossible for the agent to perform the act in question. At this point we need to consider the scope of rational possibility.

³⁵ The matter is slightly complicated by appropriately distinguishing between *existing* (or *real*) and *actual* acts, states of affairs etc. On this distinction, satyrs, hobgoblins, fictional characters and objects of false beliefs are non-existent, but contingently actual. The limitation discussed here is one on both the existence and actuality of an act and reason relation (except in the sense in which reason relations may be actual, in virtue of being an object of a false belief). Also see n9 in this chapter.

B. Non-Rational Intervention

John McDowell argues that an agent could come to believe an external reason statement and be properly motivated by it (without this meaning that the reason statement is in fact an internal reason statement) where the transition to believing the external reason statement is effected by a *non-rational* process.³⁶ A non-rational process of the kind McDowell envisages includes: having a good upbringing, hypnosis, inspiration, conversion and moving rhetoric. McDowell argues that such a process may in fact be necessary for someone because of the very fact that she has not had a good upbringing and partly due to this she cannot come to see that she has reason not to, for example, assassinate her mark for financial gain. If non-rational intervention would result in the agent coming to believe a reason statement she would not have otherwise believed, then McDowell considers that she is not coming to believe the statement on the basis of 'correct deliberation' and so the statement must be a purported external reason statement. But, for all that, if the agent can come to believe an external reason statement via non-rational means, McDowell maintains that he sees no reason to deny that the relevant external reason obtains.

Williams replies that McDowell's objection goes astray insofar as Williams' necessary condition does not require that A herself could come to believe the reason statement and be able to be motivated by it. Rather, the condition only requires this of A_{BR+} .³⁷ This reply is contrary to one of Williams'

³⁶ McDowell, 'Might There Be External Reasons?', 72–4.

³⁷ Alvin Goldman uses an analogous argument against epistemic internalism. Goldman argues that it does not seem that there is justification for thinking that the agent's knowledge of justifiers must be direct knowledge (i.e. knowable by introspection, reflection or the like) as opposed to indirect knowledge (whatever this is). Knowledge acquired by the latter route still appears to be knowledge. Yet, he also explains that internalism relies on this qualification because otherwise external facts may be readily knowable and they may be justifiers. Goldman, 'Internalism Exposed', 212, 222.

alternative versions of The Explanatory Constraint,³⁸ but, nevertheless, McDowell's objection applies to Modest Internalism. In circumstances where *A* either could not see a reason to Φ or could not be motivated to Φ for that reason, McDowell may claim that *A* may nevertheless have a justifying reason to Φ if she could come to see this justifying reason via non-rational intervention.

It is not clear whether McDowell conceives of non-rational intervention as merely leading *A* to a point where *A* can engage in a *further* process of correct, rational deliberation or whether McDowell conceives of non-rational intervention as entirely replacing rational deliberation. His paper is consistent with both interpretations.³⁹ If McDowell conceives of non-rational intervention as leading *A* to a point where *A* can engage in a further process of deliberating correctly, then McDowell's view may be consistent with Modest Internalism in any event. This is because on this interpretation of McDowell, it is still the case that *A* must deliberate correctly. (On one interpretation of part of Williams' reply to McDowell, this is how Williams himself responds).⁴⁰ If, for example, Grendel has grown up lighting a candle and praying on the last day of every month, it may be that if Grendel retains his faith, then when he is, say, 20 years old, he has a reason to light a candle and pray at the end of the coming month. In this case, non-rational

³⁸ Compare Williams' version of The Explanatory Constraint at: Williams, 'Internal Reasons and the Obscurity of Blame', 39[1].

³⁹ This is illustrated in the following passage, especially the phrase I emphasize: 'But there is no implication, as in Williams' argument, that there must be a deliberative or rational procedure that would lead anyone from not being so motivated to being so motivated [i.e. from not being motivated by an external reason statement to being so motivated]. On the contrary, the transition to being so motivated is a transition to deliberating correctly, not one effected *by* deliberating correctly; effecting the transition may need some non-rational alteration like conversion.' McDowell, 'Might There Be External Reasons?', 78 (bracketed insertion is mine).

⁴⁰ Williams' overall response to McDowell is that he does not think he was assuming anything that McDowell objects to. In particular, Williams explains that he did not intend to say that *A* would have to come to believe the statement 'If *A* deliberated correctly, he would be motivated to ϕ ' through deliberation. Williams, 'Replies', 187[4]–188[1].

intervention has influenced Grendel such that he may now have a justifying reason to light a candle and pray at the end of the month which he would not have had if it were not for his parents' raising him so. Yet for all this, Grendel's decision to light a candle and pray or not do so is still one that he determines himself. There is then no problem classifying Grendel's justifying reason as an internal reason.

Alternatively, if McDowell is interpreted as conceiving of non-rational intervention as entirely replacing any need for 'deliberating correctly', or at least the kind of deliberation we inherit from Williams, we appear to have an objection to Modest Internalism. We know that Williams' methods of reasoning are concerned with the process of coming to believe a reason statement. For example, we think of alternative solutions to a problem *to see* if we have a justifying reason to perform some other act that we have not yet thought of which will satisfy the end we have in mind. We think of whether there is a way we can satisfy multiple elements in our S in order to see if we have a justifying reason to perform some act that will allow us to achieve this. Yet, on this interpretation, the process of coming to believe a (purportedly pure external) reason statement is precisely the rational process that McDowell expressly asks whether it can be effected by non-rational means.⁴¹ For example, where Céline is hypnotized into performing an act that she would not otherwise do, there is no point at which she deliberates about and decides to perform the act at all. In contrast then to the first interpretation of McDowell's objection, this interpretation results in a challenge to Modest Internalism.⁴²

⁴¹ McDowell, 'Might There Be External Reasons?', 72.

⁴² The distinction is actually more complicated than so far presented in light of cases where, for example, one decides to: follow their instincts, associate with certain types of people, attend certain kinds of events, ask for advice. However, as we shall see, I need not delimit this distinction in greater detail. For further discussion see: Skorupski, *The Domain of Reasons*, 300.

However, interpreting McDowell in this way also has the appearance of undermining a distinction McDowell himself makes between 'a transition *to* deliberating correctly' and a transition 'effected *by* deliberating correctly'. For, if the clause 'a transition *to* deliberating correctly' is *not* interpreted according to our initial interpretation of McDowell as 'leading *A* to a point where *A* can engage in a *further* process of deliberating correctly' it must mean 'a transition to a position where *A* *has* deliberated correctly'. However, as McDowell supposes that the transition to being motivated by an external reason statement is effected by non-rational means, the result on this interpretation is that McDowell considers that one can reach a position in which one has deliberated correctly by not deliberating correctly. This proposition, as we can see, has the appearance of self-contradiction.

This result is avoided upon understanding McDowell's use of 'deliberating correctly' as *including* cases where the agent comes to believe a reason statement via both non-rational and rational means.⁴³ The thought is that when *A* is, for example, inspired by moving rhetoric to Φ , she undergoes a mental process starting from her *S* and leading to her coming to believe she has a reason to Φ . While she would not have come to believe that she has this reason if it were not for the inspiring rhetoric of another, she comes to believe it all the same and she does so (in this more inclusive sense) rationally.

The pseudo-problem that McDowell then faces is that, on this understanding, an agent's coming to believe a reason statement via non-rational intervention means that she is also coming to believe the reason-statement 'rationally' and hence the reason corresponding to the reason statement is an internal reason according to Modest Internalism and WIR. It

⁴³ This reading is consistent with Williams' first two papers, but so is the interpretation of 'deliberating correctly' that excludes coming to believe the reason statement by non-rational intervention.

is a pseudo-problem because McDowell's concern lies not so much in disproving the truth of Reason/ Motivation Internalism as it does in allowing for the possibility that *A* has a reason to Φ even where *A* cannot rationally come to be motivated to Φ for this reason, where by 'rationally' I mean the narrower sense which does not include 'non-rational intervention'. Thus, reasons that an agent could only believe and respond to via non-rational intervention can be accommodated by Modest Internalism and McDowell's concern does not raise a problem for it. Modest Internalism is consistent with *A* being motivated to Φ in *C* for the reason in question as a result of non-rational intervention.

None of this explains why Modest Internalism *should* employ this broader, more inclusive sense of 'rational' possibility, but I need not adjudicate between the two interpretations of 'rational'. Both result in a theory of Modest Internalism opposed to a corresponding version of Immodest Externalism and because the inclusive sense simplifies the argument below, I adopt this usage in delimiting Modest Internalism. If there is a compelling argument for the narrower interpretation, it can be added to the argument I give below to defend the narrower version of Modest Internalism.⁴⁴

The inclusive definition of 'rational' that allows for an agent to come to believe a reason statement by non-rational intervention means that cases of rational *impossibility* are analogous to cases where we recognize physical and epistemic limitations on ascribing justifying reasons to act. Compare:

⁴⁴ This broader sense of Reason/ Motivation Internalism could be framed as a theory asserting that: *necessarily*, at least one of four relations obtain between an agent's justifying reason to act and her motivational capacity. A relation arising via rational deliberation (in regard to recognizing the reason or being motivated to act on it) and a relation via non-rational intervention (in regard to recognizing the reason or being motivated to act on it).

Square: it is logically impossible for *A* to square the triangle because of the non-contingent fact that the two, different shapes will always exist.

Jump: it is physically impossible for *A* to jump a skyscraper in one leap because of the contingent fact that she has physical limitations common to all humans.

Lupus: it is rationally impossible for *A*, an accountant with no scientific knowledge, to cure her friend's Systemic Lupus Erythematosus in July 2012 because of the contingent fact that there is presently no cure for it and that *A* will not, in this instance we suppose, receive the kind of non-rational intervention that would allow her to see and act on such a reason.

Alien: it is rationally impossible for *J*, a rational alien being, to respect the welfare of a human being whom *J* encounters alone because of the contingent fact that *J* and *J*'s species believe that humanoid creatures are noxious and *J* will not, in this instance we suppose, receive the kind of non-rational intervention that would allow *J* to see and act on such a reason.⁴⁵

Ex hypothesi, in Lupus and Alien, *A* is denied the possibility of both rational action and non-rational intervention and, as a result, the relevant act in each case is deterministically impossible. The corollary is that, just as in Square and Jump, if the act in question cannot be performed at *t*, there is no act for the putative favouring relation to favour. Consequently, there can be no favouring relation and no reason for *A* to perform that act.

Furthermore, in regard to the question of whether there is a justifying reason for *A*, none of the above cases are relevantly disanalogous. In each case it is *physically* impossible for *A* to recognize and act on the relevant reason. In Square it is physically impossible in the deterministic sense for *A* to act at *t*, in virtue of the act's logical impossibility. In Jump the act is physically impossible because it defies the laws of physics and in Lupus and Alien the act is deterministically impossible. In Lupus and Alien, deterministic

⁴⁵ This hypothetical is ideal precisely because the rational agent in question is *not* human. For different hypotheticals, see: Harman, 'Moral Relativism Defended', 5.

impossibility is also relativized to a certain rational agent, act, set of circumstances and time. However, this feature does not justify distinguishing Lupus and Alien from Jump because if, hypothetically, the law of gravity only applied to A at t , we would still deny that A has a reason to jump a skyscraper in one leap at t .

Instead then, it might be argued that in Square and Jump, the agent's act is *non-contingently* physically impossible in the actual world because it defies laws of logic and physics respectively for A to Φ at t . In contrast, in Lupus and Alien, it might not have been deterministically impossible for A to Φ at t in the actual world, and hence, one may argue that in these cases it is only contingently physically impossible in the actual world for A to Φ in t . One consideration potentially driving this response is, once again, a concern for holding A responsible for Φ -ing or not Φ -ing. The underlying thought may be that, in the actual world, A could never be responsible for jumping a skyscraper in one leap, but perhaps she could be responsible for performing an act that is deterministically impossible for her to perform at t , but does not defy the laws of logic or physics. I discuss responsibility in Chapter 6. Right now, I need only recognize that there can be cases where it has *always* been deterministically impossible for A to Φ at t , as it could be supposed in Lupus and Alien. In such cases, holding A responsible is not a relevant consideration.

A second consideration potentially driving this concern is that, whereas *laws* of logic and physics ensure that it is non-contingently physically impossible for A to Φ at t in Square and Jump, nothing of this kind ensures that it was always deterministically impossible for A to Φ at t in Lupus and Alien. But this is also not true. If Lupus and Alien were cases in which it was always deterministically impossible for A to Φ at t , we would not be discussing them. Hence, while it could have been the case that 'at one point in time it was deterministically *possible* for A to Φ in t in Lupus or Alien', it is non-

contingently the case that 'the *case* we are focussing on is one in which it was always deterministically impossible for A to Φ at t' , compared to a case in which this is not true of A . The source of the non-contingency of physical impossibility differs, but this is immaterial. It is contingently the case that the actual world has the laws of physics it has, although *insofar as* we are concerned with the actual world, the laws of physics are as they are non-contingently. Similarly, we could have chosen to focus on cases resembling Lupus and Alien wherein it was *not* always deterministically impossible for A to Φ in C at t , although, *insofar as* we are only concerned with such cases, it is non-contingently true that, in a given case we are focussing on, it was always deterministically impossible for A to Φ at t .

In virtue of this analogy, we can also see why I maintain that cognitive possibility is not the only necessary condition on justifying reasons.⁴⁶ Consider the following hypothetical. A can recognize a certain consideration as giving rise to a justifying reason to save an innocent person. Yet, unbeknown to her, a microchip in her brain makes it impossible for her to be motivated to save innocent people. In such circumstances, it appears unjustifiable to maintain that: if A could not recognize this reason, she is not morally responsible for omitting to save the person in question, yet, if she could recognize the reason, and could not be motivated to act on it (due to the microchip), she is nonetheless so responsible. So long as cognitive possibility alone is a necessary condition on justifying reasons, such a result is permitted. Yet since this is implausible, rationally possibility is required.

With these differences accounted for, in each of the four cases, it is and always has been true that A could not Φ in C at t . Accordingly, if A does not have a reason to Φ at t in Square, A does not have a reason to Φ at t in Lupus and Alien. Yet, this is just to say that Modest Internalism is true.

⁴⁶ Cf. John Skorupski's discussion of WIR on this point: Skorupski, 'Internal Reasons and the Scope of Blame', 96–7.

C. Triviality

There is one final concern for Modest Internalism interpreted as admitting cases of non-rational intervention that is always lurking in the shadows of internalist theories generally (as well as response-dependent theories and any theory claiming that one 'thing' entails another). In response to a purported externalist counterexample to any internalist theory, in principle, an internalist can always reply that the relevant necessary relation is defeasible and holds except in situations of that kind. Yet, with every case admitted as an exception, the internalist theory faces the growing danger that the relevant necessary relation has become trivial in virtue of asserting:

- $\psi \rightarrow (\lambda, \text{ unless } \neg \lambda)$

In this context, Modest Internalism might be accused of being trivial by including non-rational intervention as a means by which an agent can come to recognize and be motivated to act on a reason. Yet the accusation is premature at this point. In the case of each of the forms of non-rational intervention McDowell identifies we can imagine an agent remaining insufficiently changed by the intervention to bring her to see a certain reason and be motivated to act for that reason. Thus, for example, a good upbringing does not ensure that 20 years later *A* is not in a state in which she cannot recognize a reason not to assassinate her mark for financial gain. We can also imagine, for example, an agent remaining stalwart in the face of attempts to inspire her or move her with heartfelt rhetoric. Indeed, the very concern we may have for *A* is her cold-heartedness and psychopathy.

At the limit, *A* could simply be given a pill without her knowledge which could induce her to come to recognize and be prepared to act for whatever reason one wanted. Or, a god may bring this to pass. Yet, even if Modest Internalism admits the possibility of such cases, this does not prove that the theory is trivial because, firstly, such cases are not common occurrences and,

as such, Modest Internalism may plausibly deny that *all* agents could be slipped a pill in *every* case in which an Immodest Externalist is intent on maintaining that there is a justifying reason for an agent to Φ . Similarly, even a Modest Internalist with faith in a god who regularly observes and reads about her god's hand in bringing about certain states of affairs may allow that 'there are instances in which a miracle is not wrought and in which an Immodest Externalist is nevertheless intent on claiming that the agent has reason to Φ '.

Secondly, the possibility of such cases of non-rational intervention does not show that the justifying reason in question is independent of A 's S . That is, it does not show that it is not a necessary condition of the existence of the reason for A to Φ in C at t that: A could Φ in C for that reason.⁴⁷ In Chapter 4 we saw that WER is usually interpreted as maintaining that at least one justifying reason is actually independent of A 's S , that is, independent of A 's S in the *actual* world. I interpreted independence in WER in this way because most, if not all, proponents of WER believe that some justifying reasons are actually independent of A 's S . Yet, an amendment to essential independence is still open to Modest Internalism and, on this understanding of independence, what *Immodest Externalists* need to show is that at least one justifying reason exists independently of A 's S , not that A could come to recognize and be motivated to Φ in C for that reason.

III Williams' Argument from Parsimony

According to Williams' second argument against pure external reasons it is not necessary to consider that there are pure external reasons because

⁴⁷ For the distinction between essential and modal independence see Chapter 1, I.

evaluations already perform this function.⁴⁸ Thus, in regard to an illustration of a neglectful husband, H, Williams opines:

There are many things I can say about or to this man: that he is ungrateful, inconsiderate, hard, sexist, nasty, selfish, brutal, and many other disadvantageous things. I shall presumably say, whatever else I say, that it would be better if he were nicer to her.⁴⁹

On the question of whether H has a pure external reason to be nicer to her, 'what', Williams asks, 'is supposed to make it appropriate, as opposed to (or in addition to) all the other things that may be said?'⁵⁰ Since ascribing a pure external reason to him does not say more than that it would be 'better' if he did not mistreat his wife, Williams accordingly considers the purported role of pure external reasons to be superfluous. We could consider the evaluative term 'better', for example, to be superfluous rather than the pure external reason but we can understand Williams to also be making the point that in saying 'that it would be better if he were nicer to her', 'better' does not carry the unwanted potential connotation of it being rationally possible for A to have acted for the reason in question. In contrast, the assertion 'there is a reason for him to be nicer to his wife' can be so understood. Accordingly, if pure external reasons do not serve a distinct purpose then it may be that the assertion that he has such a reason is false or meaningless because there are no pure external reasons.⁵¹

⁴⁸ Williams does not talk about pure external reasons, only external reasons but I omit this qualification in what follows.

⁴⁹ Williams, 'Internal Reasons and the Obscurity of Blame', 39[4].

⁵⁰ Id.

⁵¹ Insofar as reasons are taken to be metaphysical entities a qualitative version of Occam's Razor also tells against a judgment that such a pure external reason exists where their function is redundant. Although, not all writers consider that Occam's Razor justifies such skepticism. For discussion see: Daly, *An Introduction to Philosophical Methods*, 142–52.

A. Reply: Evaluations Entail Reasons

This argument (hereafter: The Argument from Parsimony) also readily supports Modest Internalism. It has however been met with the response, by more than one writer, that the application of evaluative concepts in this way is in any event reason-implicating or that it entails reasons.⁵² Indeed, in another context Williams also notes that thick concepts are ‘characteristically related to [pro tanto] reasons for action’, although what reason is provided and to whom depends on the situation.⁵³ Furthermore, Williams later acknowledges that The Argument from Parsimony is not a knock-down argument⁵⁴ against WIR and he *accepts* that there is a sense in which there is a reason for A_{BR+} to Φ even where it is not (rationally) possible for A_{BR+} to have acted otherwise. In regard to the inconsiderate, cruel husband, Williams writes:

I agree that the agent’s faults can be understood in terms of a failure to see certain considerations as reasons, just as the opposed virtues can be understood as dispositions to see those considerations as reasons. I also agree that if we think of this as a deficiency or fault of this man, then we must think that in some sense these reasons apply to him; certainly he cannot head off the criticism by saying that the reasons do not apply to him because he does not have that kind of *S*, as someone else might appropriately say that the fact that a brilliant new opera is being staged in New York is not a reason for him to go there, because a taste for opera is no part of his *S*. This is a point about the (special kind of) universality of (this kind of) reasons.

But none of this implies that these considerations are already the defective agent’s reasons: indeed, the problem is precisely that they are not.⁵⁵

⁵² Scanlon, *What We Owe to Each Other*, 367; Hooker and Streumer, ‘Procedural and Substantive Practical Rationality’, 71; Bedke, ‘Practical Reasons, Practical Rationality, Practical Wisdom’, 107; Gibbard, ‘Reasons Thin and Thick’, 295. Moreover, it is to Thomas Scanlon’s version of this argument that Williams acknowledges a ‘sense’ in which there is a reason for A_{BR+} to Φ even where A_{BR+} cannot Φ .

⁵³ Williams, *Ethics and the Limits of Philosophy*, 140.

⁵⁴ Williams, ‘Postscript: Some Further Notes on Internal and External Reasons’, 95[3].

⁵⁵ *Ibid.*, 96.

Williams accordingly allows that reasons ‘apply’ to this agent in some sense, though, in some other sense, they are not ‘already’ his reasons. This point follows, he notes, from the principle of the universality of justifying reasons, a principle claiming that:

if agent A_1 has a justifying reason to Φ in circumstances C at t , then agent A_2 would have a justifying reason to Φ in C at t .⁵⁶

Ordinarily, the level of detail in which C is specified is not held fixed. It can be relatively course-grained or not and it can take note of subjective characteristics of agents or not. The principle itself is accordingly neutral on the issue of whether all justifying reasons have some kind of internal relation to agents’ subjective conditions. In this regard, Thomas Scanlon explains:

Even if all reasons are based on desires the principle still holds that if I have a reason to do something because it will satisfy my desire, then anyone else who has that same desire and whose situation is like mine in other respects also has this reason.⁵⁷

However, both the context in which Williams uses this principle and Williams’ parenthetical remarks suggest that Williams considers that moral reasons ‘apply’ to individuals independently of their S . Yet if Williams is distinguishing between two senses of ‘there being a reason for an agent’ (or ‘that agent having a reason’), his allowance that there is a sense of any kind of a reason to Φ in the case of H concedes the point from WER that A does have a reason. In the passage in which Williams initially describes H , he clearly stipulates that H ‘really is a hard case’ and that ‘there is *nothing* in his motivational set that gives him a reason to be nicer to his wife as things are’.⁵⁸ So it is clear that Williams does not think that H perhaps could Φ or that we do not know if he could Φ . Contrary to WIR, Williams thinks that H

⁵⁶ The principle is articulated first, so far as I know, by Thomas Scanlon. Scanlon, *What We Owe to Each Other*, 73–4. See also: Skorupski, ‘Irrealist Cognitivism’, 440–3.

⁵⁷ Scanlon, *What We Owe to Each Other*, 74.

⁵⁸ Williams, ‘Internal Reasons and the Obscurity of Blame’, 39[3].

cannot rationally Φ (or that his A_{BR+} cannot rationally Φ) and yet Williams nevertheless considers that there is a sense in which there are reasons for H to Φ .

In this regard Williams errs: no such qualification is required. In the quotation above, Williams explains that ‘certainly he cannot head off the criticism by saying that the reasons do not apply to him because he does not have that kind of S' . I discuss an objection to Modest Internalism deriving from the need to hold agents responsible in Chapter 6. What I respond to here is the reply to The Argument from Parsimony that evaluations entail reasons for action so that the evaluation that H is cruel entails a pure external reason for him to be nicer to his wife.⁵⁹ This reply is potentially of concern because, if successful, it not only answers The Argument from Parsimony, it provides an argument for the existence of external reasons via evaluations

B. Evaluative Judgments & Reasons

There are two possible motivations for the objection that evaluations entail reasons. The naïve motivation is that one wants to say that ‘It is true that A has a reason to Φ ’. As we shall see, this response is ineffective because the truth of A having a reason to Φ is what is in issue.

The more sophisticated motivation driving this objection is that an agent’s evaluation of H’s conduct as cruel (irrespective of whether H can see reason to do otherwise (Φ)) provides *evidence* that H has an external reason to Φ and hence that external reasons generally exist. In this regard, I agree that we use evaluative terms that might be described as being ‘thick’ in virtue of necessarily bearing both descriptive and evaluative components, though I

⁵⁹ This is a version of buck-passers’ *negative* thesis, one that does not affect their positive thesis. See: Schroeder, ‘Buck-Passers’ Negative Thesis’.

have doubts that there are thick evaluative *concepts*.⁶⁰ Nevertheless, I leave these doubts aside.

I also agree that evaluations at least entail *pro tanto* subjective reasons for action. Yet evaluations involving thick evaluative terms only embody or entail reasons to act for agents disposed to use the evaluative concept in the same way. Thus, if *A* judges that owning up to her mistake is honourable, then she will have a reason to own up to her mistake. In contrast, if *H* does not and (rationally) cannot consider his conduct to be cruel, ignoble or otherwise undesirable then *H* does not have a reason to be nicer to his wife in virtue of such an evaluation from his own perspective. In addition, since *H* rationally cannot consider his conduct to be cruel, in this hypothetical, he also does not have a reason to be nicer to his wife in virtue of an *observer* judging that his conduct is cruel (whether or not the observer also genuinely judges that *H* cannot recognize that his conduct is cruel or that *H* otherwise has a reason to be nicer to his wife). As mentioned when discussing external reason statements, it will be a rare case where an observer knows with certainty or a high degree of confidence that it is rationally impossible for *H*, for example, to recognize that his conduct is cruel or that he (*H*) otherwise has a moral reason not to conduct himself in that way. This is partly because in most cases it would be rationally possible for *H* to recognize this and partly because of the inherent difficulty in knowing with a high degree of confidence that it is rationally impossible for someone to recognize a reason not to do something that most people would consider to be cruel. Yet, for all this, where an observer does know with a high degree of confidence that *H* could not recognize his conduct as being cruel and also judges *H*'s conduct

⁶⁰ My concern is the familiar worry, perhaps expressed by Aristotle and Kant, that thick evaluative concepts require evaluation before they can be used in their ordinary sense: Aristotle, *The Nicomachean Ethics*, bk. II.vii, pp.43–6, 1107a30–1108b10; Kant, *Groundwork for the Metaphysics of Morals*, sec. 1. This is because we can imagine circumstances in which being courageous is not favourable and in which being cruel is. See also: Blackburn, *Ruling Passions*, 94–7. Cf. Skorupski, *The Domain of Reasons*, secs. 4.5–4.6.

to be cruel, I doubt that the observer would judge that H has a reason to omit being so neglectful simply because, in this hypothetical, the observer knows that H could not recognize that his conduct is cruel or that he otherwise has a reason to be nicer to his wife. In this case, the observer's evaluation instead merely provides a reason to act for everyone disposed to use the concept of cruelty in the same way as the observer. This means that the observer's evaluation that H's conduct is cruel provides a reason to act for those agents who understand the term cruelty in such a way that they would also be disposed to judge H's conduct as cruel. In addition, it can be said that, relative to H's S, H has a reason to take certain steps to *come to see* that an evaluation of cruelty applies to his conduct so that this evaluation will provide him with a reason to act in a certain way in the future.

C. Gibbard's Objection

Allan Gibbard denies the plausibility of maintaining that judging that someone's conduct is cruel only provides a reason to act for people disposed to understand the concept in the same way. Gibbard imagines two proponents of Reason/ Motivation Internalism who stand ready to use the thick evaluative term 'cruel', though they disagree on whether or not a particular type of case is cruel.⁶¹ I provide a different illustration to invigorate discussion. Suppose that Tom considers that non-consensual retinal scanning in banks and airports is disrespectful and because of this he has reason to oppose it (Φ). Stacy in contrast does not consider retinal scanning in such circumstances to be disrespectful, but rather important for public safety, and she therefore has no reason to Φ , oppose it. Furthermore, each agent knows that the other is a proponent of Reason/ Motivation Internalism. Thus, Tom knows that Stacy is a proponent of Reason/ Motivation Internalism and so Tom knows that Stacy does not have a reason

⁶¹ Gibbard, 'Reasons Thin and Thick', 303–4.

to Φ because Stacy does not regard non-consensual retinal scanning as disrespectful. Conversely, Stacy knows that Tom is a proponent of Reason/Motivation Internalism, and so Stacy knows that Tom does have a reason to Φ because Tom regards non-consensual retinal scanning as disrespectful. Furthermore, they each agree on certain facts:

1. that retinal scanning (as well as digital fingerprinting, digital imaging of body parts and other forms of biometric identification in general) identifies individuals directly and therefore far more accurately than photographs, social security numbers, passports, identification cards, and other measures;
2. that retinal scanning is more efficient and sometimes irreversible in its capacity to identify people;
3. That there is a risk that information acquired is misused;
4. That retinal scanning therefore 'takes a piece of us' in a way that threatens our need for privacy, our sense of self, our bodily integrity, our autonomy, our dignity and security that less robust security measures do not; and
5. That its non-consensual use in, for example, international sporting events, banks, concerts, and airports may enhance public safety.

At this point Gibbard would want to state that it is unclear what is left in dispute for Tom and Stacy when the issue of whether retinal scanning is disrespectful to personal privacy arises.⁶² That is, how can Tom carry on a disagreement with Stacy about whether they have a reason to oppose the use of such technology when Tom knows that, as a proponent of Reason/Motivation Internalism, Stacy will only have a reason to oppose it if she thinks it is disrespectful. Since she does not, and since Stacy already agrees on the non-evaluative facts of the case, Gibbard does not know what Tom could possibly say. In contrast, Gibbard observes, a person who is opposed to Reason/Motivation Internalism can say that what is left for Tom and Stacy

⁶² Id.

to dispute is whether or not to oppose non-consensual retinal scanning at banks and airports (and hence whether or not they have a reason to Φ).

Gibbard's objection may be answered in one of two ways: either by showing that there actually is no disagreement between Tom and Stacy or by explaining what their disagreement consists in. Both answers are open to endorsement but I only commit to the latter answer here.

It may firstly be said that if Tom and Stacy were Modest Internalists, then it would not be so simple for each of them to determine whether the other had a reason to endorse non-consensual retinal scanning or not. This is because according to Modest Internalism, whether or not Tom and Stacy each have a reason to Φ is dependent on whether *Tom+* could recognize this reason, where *Tom+*, in contrast to Tom, reasons correctly. The same point applies to Stacy. Thus, before we can say that, as Modest Internalists, Tom and Stacy have a disagreement, it must be determined that *Tom+* and *Stacy+* would disagree and there is no assurance that this will be so. It was already stipulated that Tom and Stacy agree about the facts but if we interpret this agreement as I think Gibbard intended, then Tom and Stacy will agree about a whole range of pertinent facts that define the proper subset of non-consensual retinal scanning cases that they disagree on.⁶³ The proper subset may be small or large, the point is just that the entire range of details delimiting the subset of relevant cases will be available for them to draw on. More importantly, it may be that if Tom and Stacy's reasoning is correct, they would come to the same decision about whether or not non-consensual retinal scanning in the subset of cases they are concerned with is disrespectful. Correct reasoning is relevant in two respects. Firstly, *Tom+*

⁶³ I say 'proper subset' because I imagine there would be clear cases of non-consensual retinal scanning which were thought to be disrespectful as when, for example, one is non-consensually monitored in one's own private residence, other things being equal.

and Stacy+'s reasoning is correct in their reflection on their respective S's, an act that may change their S's. Secondly, their reasoning is correct insofar as they reason from their revised S's to their conclusion of whether or not the subset of cases they are considering are disrespectful or not. So clarified, it may be that Tom+ and Stacy+ would actually agree on the matter.

If they did not, due to differences in their respective S's, they may discuss their disagreement about whether the given case should be considered disrespectful or not. This might, for example, consist in considering cases which they both agree are disrespectful with a view to reasoning (correctly) that the subset of cases in question is relevantly analogous or disanalogous. If agreement is reached in this way then, either Tom+ will lose his reason to Φ or Stacy+ will acquire a reason to Φ .

If it is thought that Tom+ and Stacy+ may still disagree on the matter, Stacy+ might nevertheless evaluate the subset of non-consensual retinal scanning cases they are considering as harmful, presumptuous, intolerant, dictatorial, unfair, dishonourable, selfish, deceptive, misleading, or unjustified (compared to *disrespectful*) or she may evaluate the cases they are considering in a number of other ways that entail that Stacy has a reason to Φ . If this is the case, then Tom+ and Stacy+'s disagreement over whether non-consensual retinal scanning is disrespectful can be understood merely as a terminological disagreement; Tom+ has a slightly different thick evaluative concept to Stacy+, albeit a concept corresponding to the same term (disrespectful). If, however, Stacy+ does not consider the subset of cases of non-consensual retinal scanning under consideration to be cases in which an evaluation applies that entails a reason to Φ , Tom+ and Stacy+ may identify a case that they both agree is *dishonourable* (or harmful, intolerant, deceptive etc.) and (correctly) work out how cases of non-consensual retinal scanning relevantly differ if at all.

Of course it may be that, for all this, one is still not satisfied that Tom+ and Stacy+ would agree on the application of *any* evaluative term that entails a reason to Φ . But if this is true, then we are now at a point where Tom+ and Stacy+'s position is no different to the Immodest Externalist's position. In disagreeing about the application of all evaluative terms that would entail a reason to Φ , ultimately what Stacy+ and Tom+ are disagreeing about is whether they have a reason to Φ and this is precisely what Gibbard identifies as what a proponent of Reason/ Motivation Externalism would consider Tom and Stacy's disagreement to consist in.

We have then come full circle. As Modest Internalism can also maintain that Tom and Stacy's disagreement in evaluating a proper subset of cases of non-consensual retinal scanning lies in whether they each have a reason to oppose non-consensual retinal scanning in certain public places, Reason/ Motivation Externalism does not have an advantage in explaining such disagreement. We can therefore iterate Williams' point that evaluations are sufficient to express our approval and disapproval of other rational agents' conduct when they (rationally) could not have acted otherwise. In light of this, it is unnecessary to posit pure external reasons.

In this chapter we have seen that both linguistic and intuitive evidence favours Modest Internalism and disfavours Immodest Externalism. We have also seen that a commitment to pure external reasons is unnecessary. At this point, however, many metaethicists' will remain unconvinced in light of certain objections to WIR. I turn to these next.

6

Modest Internalism II

CAMILLO Be advised.

FLORIZEL

I am, and by my fancy. If my reason
Will thereto be obedient, I have reason;
If not, my senses, better pleased with madness,
Do bid it welcome.

—Shakespeare, *The Winter's Tale*, IV.4

Chapters 4 and 5 considered Reason/ Motivation Internalism (P_4 in The Relative Reasons Argument) and three arguments in favour of one particular version of this, Modest Internalism:

Modest Internalism: There is a sufficient reason or conclusive reason, R , for A to Φ in C at $t \rightarrow A^+$ could recognize R and A^+ could be motivated to Φ in C at t for R .

Where, necessarily, A^+ actually reasons as well as A could in C at t .
Conversely, Immodest Externalism maintains:

Immodest Externalism: \neg Modest Internalism.

In this chapter, I defend Modest Internalism against several objections. There are at least 21 different objections in the literature put to either Reason/ Motivation Internalism generally, or WIR, and at least five further writers

contend that either the distinction between internal and external reasons is false or that another distinction is more relevant.¹

We have already considered and dismissed six objections. From Chapter 4 we have seen that Modest Internalism avoids the charge that it cannot explain cases of *akrasia* and *accidie* because the theory merely maintains that *A*+ *could* recognize and could be motivated to act on the reason in question. We also saw that Modest Internalism avoids the conditional fallacy and cases of wayward causation in Robert N. Johnson's trilemma.

In Chapter 5 we observed that whether or not Williams' version of The Explanatory Argument begs the question, the defence of Modest Internalism does not. We also determined that linguistic evidence does not support Reason/ Motivation Externalism, thereby dispelling a linguistic argument for Immodest Externalism and the derivative objection that Modest Internalism has the burden of proving that there are only internal reasons. Finally, Modest Internalism is not vulnerable to John McDowell's concern that one may have a reason to Φ as a result of non-rational intervention: Modest Internalism can accommodate this without being trivial.

Of the remaining objections, some could be dealt with fairly swiftly either because the argument misinterprets WIR,² otherwise does not apply to

¹ David Velleman, Joshua Gert and David Wong argue that the distinction is false. Velleman, *The Possibility of Practical Reason*; Gert, 'Skepticism about Practical Reasons Internalism'; Gert, 'Internalism and Different Kinds of Reasons'; Gert, 'A Functional Role Analysis of Reasons'; Wong, 'Moral Reasons: Internal and External'. Stephen Finlay argues that distinction is real but insignificant and that it is an alternative distinction that is important. Finlay, 'The Reasons That Matter'.

² This is true of the argument given by Derek Parfit and Richard Joyce that Williams assumes the truth of The Humean Theory of Motivation. Parfit, 'Reasons and Motivation', 112, 115; Joyce, *The Myth of Morality*. As John Skorupski and Stephen Finlay recognize, Williams does not make this assumption: Skorupski, 'Internal Reasons and the Scope of Blame', 82; Skorupski, *The Domain of Reasons*, 250–1; Finlay, 'The Obscurity of Internal Reasons', 9 cols. 1–2.

Modest Internalism,³ or is not so threatening even if sound.⁴ Other objections require more discussion and I examine the most significant objections here. However I do not have space to address all remaining objections⁵ so I must instead rely on arguments from Chapter 5 in favour of Modest Internalism as well as the basis they provide for the pattern of responses to objections presented below.

It is also true of the objection, endorsed by several writers, that WIR fails to distinguish explanatory and justifying reasons: Bond, *Reason and Value*, 27–8; Blackburn, *Ruling Passions*, 264–5; Shafer-Landau, *Moral Realism*, 178; Crisp, *Reasons and the Good*, 43; Bedke, 'Practical Reasons, Practical Rationality, Practical Wisdom', 103; Joyce, *The Myth of Morality*, 108–9. (Joyce, however, then defends WIR).

Contrary to this, as noted, Williams regards a 'sound deliberative route' as correcting errors of facts and reasoning partly because internal reasons are not solely concerned with explanation, but also with the agent's 'rationality'. Williams also states the importance of interpreting an internal reason statement with '*normative force*'. Williams, 'Internal Reasons and the Obscurity of Blame', 36[2] (emphasis present). He is reluctant to distinguish the two types of reasons, noting that they are 'closely involved' and that we can explain *A*'s conduct in terms of his reasons for doing that thing. Yet this does not mean that *A* did not also have justifying reasons for acts that he could have, but *did not*, perform. Williams, 'Postscript: Some Further Notes on Internal and External Reasons', 93[1].

Ultimately then, James Lenman is right: proponents of internalism appreciate the distinction but question whether we could retain any real purchase on what our talk of justifying reasons is about if justifying reasons had so little to do with what explains action: Lenman, 'Reasons for Action: Justification vs. Explanation', sec. 3; Indeed, Gilbert Harman answered this objection *before* Williams' original paper: Harman, 'What Is Moral Relativism?', 153.

³ For example, Mark Schroeder argues that Reason/ Motivation Internalism cannot accommodate the intuition that some pro tanto reasons are pure external reasons as, for instance, where Joel has a pro tanto reason to move to Wisconsin because it has chocolate-cayenne-cinnamon flavoured ice-cream, although he has many weighty reasons to stay where he is: Schroeder, *Slaves of the Passions*, 166. Since Modest Internalism only applies to sufficient and conclusive justifying reasons, it slips past this objection. Chapter 4, III explains why insufficient reasons are precluded from Modest Internalism.

⁴ This is, for example, true of David Velleman's and Joshua Gert's arguments that the distinction between internal and external reasons is false. Velleman, 'The Possibility of Practical Reason'; Gert, 'Skepticism about Practical Reasons Internalism'; Gert, 'Internalism and Different Kinds of Reasons'; Gert, 'A Functional Role Analysis of Reasons'.

⁵ I leave aside arguments from: Haines, 'Internalism and Moral Training'; Parfit, 'Rationality and Reasons'; Scanlon, *What We Owe to Each Other*; Wong, 'Moral Reasons: Internal and External'; Arguments based on the over-generation and under-generation of reasons briefly discussed in: Finlay and Schroeder, 'Reasons for Action: Internal vs. External'. Alan H. Goldman responds to Parfit's worry in his: Goldman, 'Desire Based Reasons and Reasons for Desires'.

This chapter proceeds in four sections. Section I briefly discusses an objection to WIR and Modest Internalism that justifying reasons may exist independently of explanatory reasons because they express a truth. In Section II, I consider various purported counterexamples to Modest Internalism and Reason/ Motivation Internalism generally. Section III reviews an argument for the existence of pure external reasons grounded in a concern for holding relevant agents responsible for their conduct. Section IV discusses an argument for pure external reasons based on an analogy to theoretical reasons. Ultimately, each is unsound.

I Justifying Reasons & Truth

One objection to WIR is that it need not be a necessary requirement on a justifying reason that the justifying reason be or possibly be an explanatory reason *at all*; justifying and explanatory reasons can exist independently of one another.⁶ This objection therefore denies The Explanatory Constraint, the first premise in The Explanatory Argument and also threatens Modest Internalism. However, insofar as this objection takes the two reasons to be independent there must be some purpose to justifying reasons, other than potential explanation. The only suggestion I know of is that its purpose is to express the truth: it is true that *A* has a reason to Φ or that *A* should Φ , whether or not this is a justifying reason *A*+ could possibly act on in some sense.⁷

⁶ Parfit, 'Reasons and Motivation'; Pettit and Smith, 'External Reasons', 143[2].

⁷ Parfit, 'Reasons and Motivation', 111; Bedke, 'Practical Reasons, Practical Rationality, Practical Wisdom', 105–6; Crisp, *Reasons and the Good*, 43; Pettit and Smith, 'External Reasons', 156. Sharon Ryan also makes this point, but it appears that she is not so much begging the question as relying on an intuition. Nevertheless, even if such an intuition is sufficiently widespread as to warrant some evidence, it cannot be determinative. Ryan, 'Doxastic Compatibilism and the Ethics of Belief', 51. David O. Brink makes an analogous argument in normative ethics: Act Consequentialism merely identifies a criterion by which moral rightness can be

If this is the cited purpose of justifying reasons then we can see that the objection of distinguishing explanatory and justifying reasons is the same as another that is sometimes independently put to WIR: that in stating that the neglectful husband has a reason to be nicer to his wife what is being asserted is just that it is *true* that there is a reason for him to be nicer; it is not necessarily being stated that this is a reason he can come to in his deliberations.⁸

While we have already seen that Williams assumes the truth of The Explanatory Constraint, Williams may nevertheless reply that this assertion above also lacks justification. Critics of WIR assert that: “in stating that ‘*A* has a reason to Φ ’ they are merely identifying a truth and not also a possible deliberative consideration” and similarly leave this assertion standing without support.⁹ I imagine their reasoning is just that *Williams* has simply overlooked the purpose of their project—to state a truth—and so Williams’ argument is unsound and no more needs to be said. But this does not suffice to show that Williams’ argument is unsound because the very issue in question is whether The Explanatory Constraint does constrain the circumstances in which it is *true* that there is a reason for an agent to act. Furthermore, the objection commits its proponents to the undesirable conclusion that there is a justifying reason for *A* to ‘run faster than a fired bullet’ or ‘cone the square and not cone the square’ in circumstances where doing so would, for example, save the shooter’s intended target. Williams’

determined: it does not claim that the agent should employ this same criterion in her deliberations (as Bernard Williams considers): Brink, ‘Utilitarian Morality and the Personal Point of View’. Indeed, this is one of the definitive points of ‘Act Consequentialism’ (and most Direct Consequentialist theories) because all such theorists deny that moral deliberation should be employed directly in this way, including its classic proponents.

⁸ This reply also coheres with their reply that: internal reasons are no less metaphysically suspect than pure external reasons. Parfit, ‘Reasons and Motivation’, 129; Bedke, ‘Practical Reasons, Practical Rationality, Practical Wisdom’, 109. Presumably, the thought here is that this is because both types of reason merely assert a truth.

⁹ See this chapter, n7.

respondents cannot arbitrarily deny The Explanatory Constraint without begging the question that they accuse Williams of begging. The truth or falsity of The Explanatory Constraint has to be acknowledged by both parties as an issue in need of explicit consideration.

II Counterexamples

Three types of counterexample in the literature could be adapted and put to Modest Internalism. The first two types can be given the same structure as Mary's case in The Knowledge Argument.¹⁰ Thus, each type can be understood as:

- 1) describing circumstances in which a rational agent cannot recognize a purported justifying reason for her to act; and
- 2) describing her subsequently leaving those circumstances and recognizing this reason.

The analogy is closer for WIR than Modest Internalism because, unlike Mary and A_{BR+} , on Modest Internalism $A+$ does not have full information. Furthermore, whereas in Mary's case it is assumed that there is a fact about what red looks like (compared to what *type* of fact it is) the issue here is the *existence* of the reason. Despite this, the structure of the counterexample is the same relative to $A+$'s knowledge base and, as we shall see, the appropriate response to each is analogous to Daniel Dennett's response to The Knowledge Argument.

¹⁰ See Chapter 2, II.

A. Type 1: Williams, Shafer-Landau, Kahane & Markovits

The first type of counterexample is the most straightforward and is exemplified in illustrations provided by Williams, Russ Shafer-Landau, Guy Kahane and Julia Markovits (Williams is an Immodest *Externalist* on the definition of Immodest Externalism, given his more expansive understanding of $A+$ as A_{BR+}). In Chapter 4 we saw that where A believes that her glass of gin contains petrol, Williams argues that it is very odd to say that A has a reason to drink a glass of petrol when he believes it to be a glass of gin. It is more natural to say he does not have this reason, though he thinks he has. For this illustration to work as Williams wants it to, three qualifications are required. Firstly, we must stipulate that, A rationally cannot come to know that the glass of gin contains petrol, in the sense that, if A reasoned as well as he could without granting him all true beliefs and no false beliefs (or 'full information'), he could not come to know this. If A rationally could know that the glass contains petrol, in this sense of rational possibility, then the illustration provides no motivation for Williams granting $A+$ full information, and also, for our purposes, it would not pose a problem for Modest Internalism. We may recall that one possible outcome that $A+$ may, and perhaps often, will reach is that he should acquire more information before making a decision. Let us then amend Williams' hypothetical by stating that while A 's glass contains gin it is poisoned and that A could recognize that it is poisoned by analysing a sample of it using certain equipment he has access to. However, A cannot identify that his gin is poisoned by smelling, tasting, feeling or observing the contents of his glass. Furthermore, A has no reason to suspect that the gin is poisoned for any other particular reason and he wants nothing more than a glass of gin. Does A have a reason to omit drinking from the glass of gin before him? The answer depends on whether he has a reason to analyse a sample of the gin. If he does, then he may very well have an internal reason to omit drinking from the glass of poisoned gin. However, if he does not, as it may be here,

and if analysing the sample is the only way he could learn that his gin is poisoned, then he *may* have no reason to omit drinking his gin. I emphasise 'may' here because, for all that has been said, *A* could come to learn that his drink is tainted via non-rational intervention. He could, for example, arbitrarily decide to suspect that *this* glass of gin could be poisoned and consequently test a sample of it, or someone could enter the room and sip his gin with some infelicitous consequences. Such possibilities resemble the hypothetical in which someone slips a rational agent a pill making her believe a certain proposition and perform a certain act. From Chapter 5 we saw that Modest Internalism may count such cases as entailing that there is an internal reason for *A* to Φ , depending on whether one interprets 'rational possibility' broadly to allow for potential non-rational intervention. To make Williams' point then, and consequently also provide a counterexample to Modest Internalism, we must stipulate that in Williams' case *A*⁺ could not rationally recognize any reason to find out more information about his circumstances and his drink, where this includes, let us say, the possibility of non-rational intervention providing him with such a reason.

We also need to take Williams' illustration as showing, not just that there is *not* a justifying reason for *A* to drink from the glass, but that, more positively, *A* *does have* a justifying reason, namely, a reason to omit drinking the contents of the glass. Finally, for our purposes we need to regard Williams' illustration as going beyond making a point about a reason *statement*, to making a point about an actual justifying reason. So understood, this hypothetical might be able to make the point that Williams needs to make: that in these circumstances it is odd to say (or for us to think) that '*A* has a reason to drink from his glass', it is more natural to say (or for us to think) that '*A* has a reason not to drink from his glass, although, he thinks he has a reason to.' If one agrees with Williams, and one further agrees that *A* actually does have a reason not to drink from his glass, then one thinks that *A* has a reason not to do something, when *A* cannot possibly recognize this

reason. Hence, Williams' hypothetical, so qualified, could hold itself out as a counterexample to Modest Internalism.

Shafer-Landau, Kahane and Markovits's counterexamples each slightly differ from one another, although it will suffice to consider Shafer-Landau's example. A person, *A*, has such 'melancholia' and 'dismal pessimism about prospects for happiness' that she anticipates not realizing pleasure from mingling and chatting with others. Yet, Shafer-Landau writes, if we suppose she did, she would find new pleasures in doing so, and consequently affirm her decision to re-engage the world. Accordingly, he thinks that there is a reason for *A* to re-engage the world, even though at the time, *A*'s motivations are such that she cannot rationally recognize this.¹¹ Again, let us suppose that this also means that, at the time in question, *A* could not find out more information that would lead her to see such a reason or otherwise receive the kind of non-rational intervention that could lead her to realize this. If there were a justifying reason in such a case, as Shafer-Landau supposes, this illustration would also be a counterexample to Modest Internalism.

In both Williams' and Shafer-Landau's illustrations, *A* would be substantially better off if she performed the respective omission and act that it is suggested she has a (pure external) reason to perform and yet, in both cases we have stipulated that she cannot rationally do so. In light of this stipulation in each case, a Modest Internalist may simply reaffirm her view, based on justification given in the previous chapter, and maintain that for each case, there is *not* the reason for *A*+ (and hence, *A*) to omit to act/ act that Williams and Shafer-Landau suppose there is. This response accordingly questions the authority of any intuition that these cases may elicit in favour of there being a justifying reason for the agent to omit to act/ act and,

¹¹ Shafer-Landau, *Moral Realism*, 185–7.

deservedly so.¹² After acknowledging that a proponent of Reason/Motivation Internalism may reaffirm her view in the face of his counterexample, Shafer-Landau emphasises the significance of the agent herself later recognizing the value of the act in question:

[T]his is not bluff, bluster, bullying, or browbeating. Here, the value of certain experiences would have been endorsed *by the agent herself*, after she has had the benefit of those experiences. There is a reason for her to extend herself, even though nothing in her present make-up suggests the value of doing so, at least partly because she herself will recognize the value that such exertions bring.¹³

Yet, a proponent of Immodest Externalism cannot have it both ways. Either in her present condition *A* cannot rationally recognize a reason to extend herself (in the broad sense of rational possibility), or she can. If she can, then she has an internal justifying reason to do so. If she cannot, then neither the fact that: **a)** she would *later* recognize the value in her having done so; nor that **b)** counterfactually, as a result of non-rational intervention she recognized the value in having extended herself, changes the fact that in her present condition she cannot rationally see a reason to do so. Without further details, the suggestion that *A*'s later endorsement of the value of her earlier self re-engaging the world potentially suffers from this confusion. If *A*'s later self does *not* know whether her earlier self could rationally see a reason to extend herself, then her assertion that her earlier self 'did have this reason' may very well be a kind of bluff. Alternatively, if *A*'s later self *does* know that her earlier self could not rationally see a reason to extend herself, then her assertion that her earlier self 'did have this reason' could be bluster, or an act of bullying or badgering herself. Thus, proponents of Modest Internalism can simply reaffirm their view: in each case, the suggested justifying reason does not exist. Shafer-Landau's agent has a reason to take certain steps that she can rationally perform and that could lead her to re-engage the world,

¹² See Chapter 5, II.

¹³ Shafer-Landau, *Moral Realism*, 186.

but there is no point asserting that she has a reason to do something beyond this.

In support of the point, proponents of Modest Internalism can iterate the problem for Immodest Externalists: if the agent rationally could not perform the act or omission for the reason in question, then to say that 's/he nevertheless has a reason to avoid drinking his gin/ extend herself' or that 'there nevertheless is a reason for him to avoid drinking his gin/ for her to extend herself' is to say that there is a justifying reason for the agent to perform an act or omission that is deterministically impossible for that agent to perform. However, we have not been able to find a relevant difference between maintaining that a justifying reason exists in a case of that kind and maintaining that, at t_1 there is a justifying reason for A to walk through a concrete wall to rescue a dying child. Unless, therefore, we consider that A has a justifying reason to traverse concrete, we should maintain that, equally, there is no reason for A to omit drinking his gin/ extend herself. Emphasising that, at a later time (t_{10}) when A has acquired the technology to walk through walls, A recognizes the value in walking through the wall to save the child at t_1 , adds no weight to the argument that he had a justifying reason to do so at t_1 .^{14/15}

¹⁴ Compare John Skorupski's defence of Cognitive Internalism against such cases: Skorupski, *The Domain of Reasons*, 248, 253–6, 394–8.

¹⁵ Kahane's counterexample involves a human agent's experiencing pain for the first time and it consciously mirrors the structure of The Knowledge Argument. However, ultimately the same response applies: the agent does not have the pure external justifying reason it is suggested that he has. Kahane, 'Feeling Pain for the Very First Time'. Julia Markovits cites a slightly different type of case to these counterexamples: an agent could act for the reason in question but would better achieve her aim if she performed the same act for a different reason. For example, a soldier fighting for peace ought not be sympathetic, but ruthless, and therefore would be better off fighting for a reason more consistent with this, perhaps the lust for war. Again, the same response applies. Markovits, 'Internal Reasons and the Motivating Intuition', 156–7.

B. Type 2: Millgram & Schroeder

In Williams and Shafer-Landau's counterexamples, it was stipulated that the relevant rational agent, *A*, could not recognize the justifying reason in question and yet it was suggested that the reason exists nonetheless. As we saw, if *A* could recognize and act on this reason in each case, then, we may agree that *A* has a reason to Φ . In contrast to this, Elijah Millgram and Mark Schroeder's counterexamples gain purchase insofar as it is true that *if A* could recognize and act for that reason, then the reason would *not* exist. Hence, if there is no external reason, then there is no reason at all.¹⁶

Millgram describes an insensitive person, Archie, whose colleagues are unpleasant to him and whose friends are as insensitive as he is. Archie's insensitivity prevents him from seeing the benefit of, for example, avoiding acquaintances in their hour of grief, or taking steps to become more sensitive. Millgram also explains that Archie cannot see reasons to act in these ways:

Archie's insensitivity is a deliberative incapacity: it consists in being unable to appreciate certain reasons for action, these among them. Because he is insensitive, he cannot see that his own insensitivity gives him reasons for action. (*'Insensitivity,'* he sniffs, his voice dripping with contempt.) If he could reason in this way (e.g., 'I had better stay away from the funeral; if I go, I'll only make things worse') he would *ipso facto* be sensitive enough not to have these reasons. Archie has a reason for action which is not such that if he deliberated correctly it would motivate him to action. For if he *could* so deliberate, then he would no longer have this reason.¹⁷

Millgram clarifies that Archie does not recognize and does not want the rewards of being more sensitive. He does not, for example, recognize and

¹⁶ These are what David Sobel calls: superfragile reasons. Sobel, 'Explanation, Internalism, and Reasons for Action', 231; Sobel, 'Reply to Robertson', 187.

¹⁷ Millgram, 'Williams' Argument Against External Reasons', 203 (emphasis present).

want the development of trust in friendships as a reward. Evaluative attitudes for such an end are not in his S. However, were Archie to experience the development of trust in friendships, 'he would acknowledge the extent to which his life had improved.'¹⁸ Millgram concludes: 'It is natural to say that Archie has reasons to change his ways, reasons which are not grounded in elements of his S and are consequently external reasons.'¹⁹

Schroeder's suggested counterexample is similar:

Nate loves successful surprise parties thrown in his honor, but can't stand unsuccessful surprise parties. If there is an unsuspected surprise party waiting for Nate in the living room, then plausibly there is a reason for Nate to go into the living room. There is certainly something that God would put in the 'pros' column in listing pros and cons of Nate's going into the living room. But it is simply impossible to motivate Nate to go into the living room for this reason—for as soon as you tell him about it, it will go away. Nate's case looks to me like a counterexample to many strong theses about the connection between reasons and motivation.²⁰

In each case, a proponent of Reason/ Motivation Internalism may again reaffirm her own version of Reason/ Motivation Internalism and maintain, as a result, that the purported reasons for Archie and Nate to act do not exist. For example, in Schroeder's illustration a proponent of Reason/ Motivation Internalism may maintain that since, *ex hypothesi*, it is impossible for Nate to go in the living room for the reason that there is a surprise party there, then there really is no justifying reason for Nate to go in there. Clearly we do not want Nate to go into the living room for the reason that there is a surprise

¹⁸ Ibid., 204.

¹⁹ Id. Joyce argues that, contrary to Millgram, if per WIR, Archie were fully informed, then Archie's S would be different and there would be the requisite relation between Archie's S and Archie changing his ways. Joyce, *The Myth of Morality*, 112–5. However, since WIR suffers from the conditional fallacy, this defence will not work in other cases, such as where, for example, Archie has a reason to find out where the funeral is being held.

²⁰ Schroeder, *Slaves of the Passions*, 165. Other writers also endorse this type of counterexample. For example: Markovits, 'Why Be an Internalist About Reasons?', 260.

party (this is the point of the illustration). Rather, the proponent of Reason/Motivation Internalism may continue, as we want Nate to experience the surprise and subsequent frivolity and mirth involved therein, we want to think that Nate has a reason to go into the living room *for some other, unrelated*, justifying reason. For example, we may hope that Nate enters the living room because his friend asks him to get him a drink, or because he left a magazine there, or for no particular reason at all (he stumbles in). But such reasons are internal reasons (at least in the broad sense of 'rational possibility' that Modest Internalism may allow, the sense that includes non-rational means). If no such alternative reason exists for Nate to enter the living room, then Modest Internalism simply maintains that there is no reason for Nate to go in.

Framed in the right way, this response need not beg the question by assuming that a justifying reason cannot exist independently of Nate's capacity to recognize and act for it. Rather, a proponent of Reason/Motivation Internalism may instead maintain that they are merely raising a doubt that the illustration favours Reason/Motivation Externalism. In raising such a doubt, one need not assume that there are only internal reasons to act.

Millgram and Schroeder provide different responses to this line of defence. Similarly to Shafer-Landau, Millgram emphasises the importance of the possibility that Archie may subsequently value the importance of his earlier self taking steps to become more sensitive to the concerns of others. Millgram then seeks to show that Archie's coming to realize this can be learned from experience itself, rather than deriving from an element that was always present in his S or his imagination. However, even if this is so, it does not establish that Archie did have the reasons it is suggested he has, at the time when he could not recognize and act on them. Moreover, even if, *per impossible*, it were true that Archie would be better off taking steps to become

more sensitive, this does not show that he had a reason to do so. Millgram begs the question by conflating this purported truth with there being a justifying reason to so act. His argument is accordingly vulnerable to the same problem as Williams', Shafer-Landau's, and Kahane's suggested counterexamples.

In contrast, Schroeder argues that the proponent of Reason/ Motivation Internalism arbitrarily reaffirms her own theory:

Of course, one way to respond to Nate's case is to say that it doesn't count, for some reason—perhaps because the fact that there is an unsuspected surprise party waiting in the living room is not really, after all, a reason for Nate to go there. But this seems to me to be *ad hoc*. If what Nate enjoyed immensely was playing poker, then the fact that there is poker being played in the living room would be a reason for Nate to go in. If what Nate really enjoyed was watching TV, then the fact that there is a TV in the living room would be a reason for him to go in. So I see no obvious reason to insist that Nate's enjoyment of successful surprise parties must be different. I see no reason other than an attachment to some theory to think that there can't be reasons that no one could ever act on. It is not as if such reasons don't matter, after all—they still play a role in determining what Nate ought to do—they still show up on God's list of pros and cons.²¹

The only way to avoid this charge is by identifying independent justification for adherence to, for example, Modest Internalism. Yet we have already seen such justification because it is largely responsible for grounding this view. Since, *ex hypothesi*, it is deterministically impossible for Nate to enter the living room for the reason that a surprise party lies in wait, there is no such act in the actual world *to be* favoured, just as there is no logically or physically impossible act in the actual world to be favoured.²² The point can be made slightly differently. It would *not* be *ad hoc* to deny that there is a reason for Nate to cone the square to save someone's life. One may instead point to the logical impossibility of the act. The same point applies to Nate entering the living room for the reason that there is a surprise party for him.

²¹ Schroeder, *Slaves of the Passions*, 165–6.

²² For the defence of this point, see Chapter 5, II.

Indeed, it is perhaps clearer in this latter case because, while an omniscient observer may think that Nate has a justifying reason to go into the living room for *some* justifying reason unrelated to there being a surprise party, the observer will not think that Nate has a reason to enter the living for the reason that there is a surprise party waiting for him. This would be to attribute to the observer a paradoxical set of beliefs.

Moreover, if Immodest Externalists agree that Nate does not have a reason to cone the square to save someone's life, as I have argued, then it now appears to be *ad hoc* to maintain that there *is* a reason for Nate to enter the living room because there is a surprise party waiting for him. Without further justification, insistence that there is this reason for Nate can only be explained by a preference to adhere to some version of Reason/ Motivation Externalism.

In sum, in illustrations of this kind, the justifying reason that it is thought that a rational agent has, may, on closer consideration actually be an internal reason bearing the same favourable consequences as the purported external reason. (For example, the justifying reason that the Immodest Externalist supposes that Nate has may really be an *internal* justifying reason to enter the living room with some object in mind, other than that there is a surprise party waiting for him). However, even where it is reaffirmed that the justifying reason in question is external, the Immodest Externalist's claim that 'a justifying reason would not exist if the agent could come to recognize it' is no bar to doubting that it exists at all. This is especially so in light of the other possible internal justifying reasons it may properly be supposed that the agent does have.

C. Type 3: Hampton & Lillehammer

The third type of counterexample also derives from the wording of a given version of Reason/ Motivation Internalism. In relation to Williams' version of Reason/ Motivation Internalism ('WIR'), Lillehammer argues that the rational process of deliberation that *A* herself ultimately employs to determine whether or not she has a reason to Φ in *C* is either defined relative to *A*'s *S* or it is not.²³ If it is, then, among other things, Lillehammer argues that this allows for the possibility that WIR endorses the view that, because of her *S*, a given agent's principles of rational deliberation include principles which cannot be consistently applied or principles which require misapplication. Lillehammer rightly considers it implausible to define 'rational process of deliberation' in a way that includes such results.

Alternatively, Lillehammer observes, if 'rational process of deliberation' is defined independently of *A*'s *S*, this entails that *A* has a reason to act in whatever way *A*_{BR+} would want her to act. It also entails that, in determining what *A* has a reason to do, *A*_{BR+} has reason to deliberate in a certain way according to the definition of 'rational process of deliberation'. Yet, Lillehammer argues, each of these reasons are reasons to act that exist independently of *A*'s *S* and so by definition they are both external reasons. Accordingly, where 'rational process of deliberation' is defined independently of *A*'s *S*, Lillehammer contends that WIR entails the existence of external reasons to act and the view is therefore self-defeating.²⁴

²³ Lillehammer, 'The Doctrine of Internal Reasons'.

²⁴ Alvin Goldman also makes this point in regard to internalism in epistemology insofar as the epistemic agent must have to choose her method of calculation and know how good her own skills are. Goldman, 'Internalism Exposed', 219–20. An analogous argument to this also arises in the literature against The Humean Theory of Motivation (HTM) which purports to show that HTM is inconsistent because the hypothetical imperative is itself a categorical imperative and Humeans must therefore endorse one categorical imperative after all. Schroeder, *Slaves of the Passions*, 41–9. Finally, John McDowell also develops such an argument against

In relation to Modest Internalism, I deny that the second kind of reason that Lillehammer identifies exists. *A*⁺ does not have a reason to deliberate in a certain way according to the definition of ‘rational process of deliberation’; *A*⁺ is a fictional entity who just does deliberate in this way. But I agree with Lillehammer that there is a sense in which the first of the two reasons that he identifies exists. Appropriately adapted for Modest Internalism, *A* does not have a reason to follow *the advice* of *A*⁺ but, *A* does have a reason to do what: **a)** *A*⁺ could recognize that she (*A*⁺) has a reason to do; and what: **b)** *A*⁺ could act on for that reason. It does not, however, follow from this that this is an external reason. Indeed it cannot because *A*⁺ is defined as being a version of *A* who reasons as well as *A* could in *C* at *t*. Ultimately then, the tenability of Lillehammer’s argument depends on whether the conditional fallacy is committed by the version of Reason/ Motivation Internalism in question. Lillehammer’s objection is sound against WIR, but ineffective against Modest Internalism.²⁵

III Holding A Responsible

One of the more significant arguments for Reason/ Motivation Externalism derives from a need to be able to hold agents responsible for their conduct and to be able to punish agents for illegal and immoral behaviour.²⁶ We want to hold serial killers responsible for their crimes and so we ordinarily

Simon Blackburn’s projectivism: a fact of the matter is required to legitimise the processing mechanism behind one’s judgment that Φ -ing is morally required/ immoral. Projectivism pictures the processing mechanism that we use to make moral judgments as ‘something that one can contemplate as an object in itself’ and ‘step back from’ and yet, McDowell argues, we cannot. McDowell, ‘Values and Secondary Qualities’, 122.

²⁵ Jean Hampton provides a similar argument: Hampton, *The Authority of Reason*, 76–7.

²⁶ See for example: Shafer-Landau, *Moral Realism*, 187.

attribute a reason to them to have done otherwise, if only a reason in a certain sense.²⁷ With this motive framing our judgment, we may certainly want to think that there *is*, at least in some sense, a reason for a serial killer to have done otherwise.

Gilbert Harman has already argued that in some cases the moral qualities of an act effect one's decision to attribute an intention to the agent in question.²⁸ In his illustration, a sniper tries to kill a person from a distance, knowing that his chances of success are slim. In such a case, Harman argues that 'the sniper does not flatly intend to kill the soldier, although, if he succeeds, he does kill him intentionally.'²⁹ In contrast where the sniper is merely targeting a bulls-eye at a firing range, Harman claims that even if it is just a lucky shot, he does not intentionally shoot the bulls-eye. Harman explains that the analysis is different in each case because 'we think that there is something wrong with killing and nothing wrong with shooting a bulls-eye.'³⁰ That is, we want to hold the sniper accountable for killing the soldier and this motivates our attribution of intention to him. But, conversely, we may not

²⁷ Williams makes a similar point: 'I suspect what are taken for external reason statements are often, in fact, optimistic internal reason statements: we launch them and hope that somewhere in the agent is some motivation that by some deliberative route might issue in the action we seek.' Williams, 'Internal Reasons and the Obscurity of Blame', 40[2].

²⁸ The point is initially made, so far as I know, by G. E. M. Anscombe and later by Gilbert Harman. Anscombe, *Intention*, §25, 44; Harman, 'Practical Reasoning', 433–4. See also: Mele, 'Action', 353–5. More recently, this point has arisen in work in experimental philosophy. Knobe, 'Intentional Action in Folk Psychology'; Knobe, 'Intentional Action and Side Effects in Ordinary Language'; Knobe, 'Intention, Intentional Action and Moral Considerations'. (In this context, the method used in experimental philosophy does not restrict the point of the conclusion because what makes this experiment interesting is not a divergence of intuitive opinion, as is often the case, but the convergence of opinion to the (perhaps surprising) result that: moral considerations can affect the attribution of intention to an agent's conduct. Accordingly, we can use the thought experiments employed in this work to support the same point here).

²⁹ Harman, 'Practical Reasoning', 433.

³⁰ *Ibid.*, 434[2].

want to reward him for being lucky (just as we may not want to reward a doctor for 'knowing' that the blood test results are negative in a Gettier case).

For the same reason it may be that where an agent's Φ -ing in C has rather undesirable consequences, in certain cases we are more likely to attribute to the agent a reason to have avoided Φ -ing in C so as to hold her responsible. If, before attributing the reason, we further turn our mind to whether it was rationally possible for A to have seen a reason to act otherwise and we have doubts about this, we will nevertheless be likely to consider that she could have acted otherwise so as to hold her responsible.

But, in doing so we are actually attributing an internal reason to the agent in question in which case Modest Internalism agrees that the agent had a justifying reason to avoid Φ -ing in C and so she should be held responsible. This means that ultimately this is a case in which what is being attributed to the agent is a mental state from which she could have performed the act or from which she in fact did perform the act in question. Accordingly, such cases need to be distinguished from cases where we know with certainty that it is not rationally possible for A to Φ in C.

A. Holding A Responsible for not Developing her S

This qualification has the potential to altogether reverse our intuition that 'there is a reason for A to Φ in C at t' '. In light of the fact that it is not rationally possible for A to Φ in C at t , the concern to nevertheless hold A accountable may be developed in two ways. On one view, it may be argued that A may be held accountable for not developing the contents of her S at some earlier time in a way that would have allowed her to recognize and respond to the appropriate reasons in C now. On an alternative view, A may just be held accountable for not Φ -ing in C now irrespective of whether she

could have developed her S in a certain way at some previous point in time. Both views are problematic.

The first seeks to hold *A* responsible for not recognizing or responding to a reason to Φ in *C* now on the ground that *A* was at some past point in time capable of developing her S in a way that would allow her to do so. It is because *A* could have Φ -ed in *C* if she put herself in a position to be able to Φ in cases similar to *C*, that we say that she has a reason to Φ in *C* now, even though she cannot Φ in *C* now. Thus, for example, we want to hold moral agents accountable for not developing the S required to have the gumption to prevent an adult harming young children.

The view may be generalized so as to apply to cases where the interests of others are not at stake (what I will for convenience call 'non-moral cases'). Thus, for example, we may want to say that Felicity has a reason to go to private social functions at the London Fashion Festival now because 10 years ago she could have chosen a significantly more enriching life as a fashion designer and met people in the industry who would have given her access to the event. In a non-moral case such as this we may (or may not) still care to hold Felicity accountable for not developing her interests in a way that would have allowed her to attend such functions.

Furthermore, the ascription of responsibility in this way need not entail that if one falls in with the wrong crowd in one's youth that one will forever thereafter be unable to see and act on the 'right' reasons. If it were to, this would undermine the point of holding agents responsible in the first instance because all agents could cite as explanation for their imperfect S, some factor from their early youth for which they are not responsible. In light of this, if agents are to be held responsible for their S, their responsibility would have to be relativized to either the present time or a certain point in their history at which they could have directed their S in the

supposedly proper way. On such a view we could qualify the necessary condition for A to have a justifying reason to Φ in C at t by providing the following alternative:

A has a reason to Φ in C at t (now) $\rightarrow A+$ could Φ in C at t (now) for that reason.

In addition, where the consequent is false, nevertheless:

A has a reason to Φ in C at t (now) *relative to* A 's S at a point in A 's history h if:

Φ -ing in C at t (now) is morally required or if Φ -ing in C at t (now) would be part of an ideal life for A (where h is a point in A 's past at which A could change herself so that $A+$ could Φ in C at t (now)).

In this way, even if right now, A cannot rationally Φ in C relative to A 's current S , A will still have (now) a pure external reason to Φ at t relative to A 's S at h .

The problem with this account is that the attribution of this pure external reason is, by hypothesis, largely, but not entirely, dependent on there being an internal justifying reason for A to develop her S in a certain way (' ψ -ing') at some earlier time, h .³¹ The entire point of holding A responsible for not ψ -ing at h instead of holding A responsible for Φ -ing at the present time (t) is that, at h A *was* rationally able to ψ . In light of this, we can only understand this 'capacity' or 'ability' of hers to ψ at h in a way that avoids the problem we have with holding her responsible for Φ -ing at t . Since the problem we

³¹ It may be thought that: it may not be that, at h , A or $A+$ could see a reason to ψ , because it may be that A or $A+$ could not see that this would lead to her having the capacity to Φ at t because t is in the future. There is much that could be said in response to this concern. Suffice it to say that, by hypothesis, in the least A has an internal reason to ψ at h via non-rational intervention for, by hypothesis, A *can* (rationally) ψ at h .

have with holding her responsible for Φ -ing at t is that it was not rationally possible for her to do this, to hold A responsible at all we must consider that it *was* rationally possible for her to ψ at h . That is, we must consider that her capacity to ψ at h is understood as rational possibility. The corollary is that we may say that A had an internal reason to develop her S in a certain way at h . But, if this is true, then we can already hold A responsible for not ψ -ing at h . A had an internal justifying reason to ψ at h , and she is responsible for the consequences of not doing so. As a result, there is no need to posit a pure external reason for A to Φ at t so as to hold A responsible for not Φ -ing at t . Perhaps we do not just want to hold A responsible for not ψ -ing at h . Perhaps we also want to hold her at least concurrently responsible for not Φ -ing at t . However, the ascription to A of a pure external reason to Φ at t is also not necessary to justify the imposition of such responsibility. The point of holding A responsible for not ψ -ing at h is just that there is a potential for circumstances to arise in which she may need to perform moral acts such as Φ -ing in C at t . Accordingly, either A 's internal reason to ψ at h or A 's internal reason to ψ at h in conjunction with the consequences of her not doing so, as they arise at t , are sufficient to justify holding A responsible for not ψ -ing at h and consequently not Φ -ing at t .³²

None of this means that Modest Internalism is committed to the claim that an agent *always* has an internal justifying reason to act or omit to act in some way that would, if she had of performed that act or omission, prevented some otherwise unforeseen disaster from occurring. Indeed, by definition Modest Internalism is only committed to the claim that ' A has a justifying reason to Φ in C at t or h if A could Φ in C at t or h '. Of course, according to Modest Internalism, it may be that A could only Φ in C at t or h in virtue of non-rational intervention. Yet, Modest Internalism only claims that this is a necessary condition for A having a justifying reason to Φ in C at t or h . Thus,

³² In this regard see Williams' proleptic account of responsibility. Williams, 'Internal Reasons and the Obscurity of Blame', 40[4]–4[3].

after graciously being invited to and attending Nate's party, if I choose to drink, I may compromise my wits in a way that prevents me from being able to help Redmond get to the emergency room. As I could have chosen not to drink, in one sense it may have been within my capacity to help Redmond. But, if I could not have known that Redmond would require my help then, I certainly could not have chosen to omit drinking for the reason that Redmond would need my help. So the attribution of an internal justifying reason to A to ψ at some point in her history h is, as always, conditional on her being able to rationally recognize and be motivated to act on that reason. Thus suppose, for example, that after almost killing someone in her latest burglary, Claire realizes that she needs to change her conduct and accordingly her life to avoid such a result in the future. We may say here that Claire has an internal justifying reason to make certain lifestyle changes and in virtue of this reason we may hold Claire responsible for the consequences of not doing so.

Thus, we can see that holding A responsible at h for not ψ -ing at h means that we need not say that A has a pure external reason to Φ in C at t . If A is already being held responsible at h for not ψ -ing at h then the whole purpose of ascribing the pure external reason to A to Φ at t becomes redundant under the theory that we can consider that A has a pure external reason to Φ so as to hold her accountable.

B. Holding A Responsible: Specific Deterrence

In contrast to the above account, it may instead be thought that we just want to hold A accountable for not Φ -ing at t irrespective of whether she could have done otherwise. That is, one may consider that there is a reason for A to Φ at t whether or not she could do so and whether or not at some stage in the past she could have developed her S to be able to do so. The rationale for imposing responsibility in such cases may lie in deterring the agent herself

from omitting to Φ in the future (specific deterrence), or in deterring other agents from omitting to Φ in the future (general deterrence) in like circumstances. Both justifications are flawed.

Where the imposition of responsibility is thought to be justified by specific deterrence, the justification lacks merit because it was not rationally possible for A to Φ at t in the first instance. It is true that the act of holding A responsible for not Φ -ing at t will nevertheless probably deter A from not Φ -ing in relevantly similar circumstances in the future. However, since A could not have done otherwise, it would be just as appropriate to hold another moral agent, B , responsible for A 's not Φ -ing so as to deter B from not Φ -ing in the future. Of course, where B is morally innocent, it is ludicrous to suggest that we should hold B responsible for not Φ -ing so as to deter B from not Φ -ing in the future. Yet, if A could not have done otherwise, then she is also morally innocent and the suggestion that we hold her responsible nonetheless to deter her is just as ludicrous.

C. Holding A Responsible: General Deterrence

Alternatively, the imposition of responsibility may be justified by the aim of deterring the population at large. In this case, the justification is consequentialist in nature as it justifies holding A responsible for Φ -ing in C (even where A could not have acted otherwise) in light of the benefits of minimizing further, similar incidents occurring. Indeed, the situation is not unlike H. J. McCloskey's sheriff who must choose whether or not to sacrifice an innocent defendant to a rioting mob to restore order.³³ In such a case it is

³³ McCloskey, 'An Examination of Restricted Utilitarianism', 468. It is also not unlike transplant cases and other situations devised to show that consequentialism has unintuitive consequences. While I do not adhere to consequentialism, for my part, I do not think consequentialism does actually entail that a certain innocent individual should be sacrificed in these situations. See, for example, Hewitt, 'Normative Qualia and a Robust Moral Realism', 286–302.

undesirable for the sheriff to sacrifice the innocent defendant and indeed consequentialists may be able to provide for this outcome. But if we agree that the innocent defendant should not be sacrificed (whether on consequentialist or non-consequentialist grounds) then we must also assert that we should not hold *A* responsible for Φ -ing in *C* so as to realize the benefits of deterring the population at large from following suit.

Furthermore, even if *A* is to be held responsible for Φ -ing in *C*, irrespective of whether it was rationally possible for her to have done otherwise, it need not follow that *A* must have had a pure external reason to Φ in *C*. Since it is agreed by all parties that it was not rationally possible for *A* to Φ in *C* it is open to hold *A* responsible for Φ -ing in *C* (if so desired) without ascribing to *A* a pure external reason to have done so. Accordingly, as the benefit of general deterrence can be realized without ascribing the relevant pure external reason, we can see that the function of pure external reasons in this context is redundant.

IV Theoretical Reasons

One underlying motivation for denying The Explanatory Constraint is the sense that there is a standard of practical reasoning that exists independently of the reasons a particular agent can recognize and be motivated by. We have seen two objections to The Explanatory Argument in Chapter 4 that may derive from this motivation: **i)** that The Explanatory Constraint is concerned with explanatory reasons rather than justifying reasons; and **ii)** that the neglectful husband has a reason to be nicer to his wife, not in the sense that he can necessarily deliberate on this reason, but just in the sense that it is true that the reason exists. We saw that both arguments are question-begging.

A more sophisticated argument deriving from this motivation may instead maintain that practical reasons may exist independently of an agent's capacity to recognize and be motivated by them just as theoretical reasons do.³⁴ This argument starts to get off the ground if we have far less doubt about the possibility of theoretical reasons of this sort existing.³⁵ For example, we may think that there is a theoretical reason for a rational agent to believe that $7 \times 13 = 91$, whether or not she could recognize this. Analogously, so the argument maintains, a given rational agent may have a practical reason to refrain from stealing a stranger's car for joyriding, whether or not she could recognize it.

The literature on theoretical and practical reasons is significant in itself. Suffice it to say here that the argument, developed thus far, is unpersuasive for three reasons. Most importantly for our purposes, the relevant analogy must be one between sufficient theoretical and sufficient practical reasons. In this regard, it must be shown that sufficient theoretical reasons cannot, or should not, be considered analogously to *sufficient practical reasons*. That is, the *direction* of the purported analogy requires justification. Secondly, it must be shown that practical reasons should not instead be understood in an analogous form to reasons to feel, or *vice versa*. Finally, it must be shown that both practical and theoretical sufficient reasons must answer this question in the same way, and, at least in the case of sufficient justifying reasons that are also conclusive, I doubt that this is so.³⁶

³⁴ Elijah Millgram makes this argument. Millgram, 'Williams' Argument Against External Reasons'. Mark Schroeder and Stephen Finlay also briefly mention it: Finlay and Schroeder, 'Reasons for Action: Internal vs. External', sec. 2.2.

³⁵ Millgram, for example, thinks that 'externalism is far more obviously false in the theoretical domain than in the practical'. Millgram, 'Williams' Argument Against External Reasons', 199. Cf. Skorupski, 'The Ontology of Reasons', 114 cols. 1–2; Skorupski, *The Domain of Reasons*, 41–2.

³⁶ Essentially, whereas conclusive *theoretical* reasons need only have the constitutive aim of truth (or true belief, or warranted belief etc.), conclusive *practical* reasons must have an additional constitutive aim insofar as *akratic* action is considered to be *irrational*. Furthermore, since an agent acts irrationally where she is *akratic*, but

V Review & Outlook

It is worth recapping the argument so far, modified in light of the expansion to The Theory of Relative Motivation (**P₅**) discussed in Chapter 3.

The Relative Reasons Argument^{II}

- P₁** A moral fact or a sincere moral judgment that an agent *A* should Φ in *C* at *t* necessarily entails a sufficient or conclusive practical justifying reason for *A* to Φ in *C* at *t*. (*Morality/Reason Internalism*)
- P₂^{*}** All sufficient and conclusive practical justifying reasons are either agent-relative or contingently agent-neutral. (*Relative Reasons^{*}*)
- P₃^{*}** According to Moral Impartiality^{*}, in propositions 1-3 of Moral Realism either:
- a)** facts exist independently of moral agents; or
 - b)** where **a)** does not apply to a given entity, any relation between the entity and moral agents is *non-contingently* agent-neutral.

-
- C₁^{*}** Moral Impartiality^{*} is not true (and hence, Moral Realism is not true).

P₂^{*} is supported by the following revised argument:

The Reason/ Motivation Argument^{II}

- P₄** All sufficient and conclusive practical justifying reasons for a rational human agent, *A*, to Φ in *C* at *t* are necessarily related

accidentally ends up acting as she initially judged she should, then this additional constitutive aim should be: **i)** Φ -ing for that reason; or **ii)** intending, deciding or planning to Φ for that reason. This additional aim makes it the case that practical reasons must be capable of being known by the agent for the reason to be a practical reason at all. Pure external practical justifying reasons might then be translated as *theoretical reasons to act*, yet reclassifying pure external reasons in this way lacks justification. For a different defence of the disanalogy see: Markovits, 'Why Be an Internalist About Reasons?', 267–70.

to A 's capacity to be motivated to intentionally Φ (or omit to Φ) in C at t . (*Reason/ Motivation Internalism*)

P₅* In the actual world and close possible worlds, rational human agents are so constituted that, it is non-contingently true that a rational human agent's motivation to intentionally Φ (or omit to Φ) in C at t for a sufficient or conclusive subjective justifying reason, is only correctly explained by:

- a) an agent-relative explanatory reason; or
- b) a contingently agent-neutral explanatory reason.
(*Relative & Contingently Neutral Explanations*)

P₂* All sufficient and conclusive practical justifying reasons are either agent-relative or contingently agent-neutral. (*Relative Reasons**)

Chapter 3 maintained that **P₅*** is false and if this is right, The Reason/ Motivation Argument^{II} is unsound. Yet, if Reason/ Motivation Internalism (**P₄**) is true and if Morality/ Reason Internalism (**P₁**) is true, certain moral facts are not independent of rational human agents, and specifically, their motivational capacities, contrary to recent robust versions of moral realism. With this in mind, Chapters 5 and 6 defended a version of **P₄**, Modest Internalism:

Modest Internalism: There is a sufficient reason or conclusive reason, R , for A to Φ in C at $t \rightarrow A+$ could recognize R and $A+$ could be motivated to Φ in C at t for R .

where $A+$ only differs from A in virtue of, necessarily, actually reasoning as well as A could in C at t . This accordingly leaves **P₁**, which we turn to next. Chapter 8 then considers the implications of this, particularly in regard to what this means for moral facts.

7

Morality/ Reason Internalism

Moral facts may be dependent on human moral agents' in more than one way. Chapter 3 maintained that for a given moral reason, all moral agents in the appropriate sub-group could be motivated to act for it in the particular circumstances and time in question. Chapters 4-6 maintained that it is a condition of the existence of a reason that, the agent whom the reason is a reason for, could both recognize and be motivated to act for the reason in question. Thus, overall, a moral reason is dependent, not just on the existence of a rational human being, but on their capacity to recognize the consideration giving rise to the reason, the reason to act itself, and their capacity to be motivated to act for this reason:

R is a justifying reason for A to Φ in C at $t \rightarrow A+$ can recognize R and be motivated to Φ in C at t for R.

I also think that this is a sufficient condition, and Section I defends this contention. In this regard it is worthwhile iterating that, as in Chapter 5, I remain neutral on the issue of whether or not the concept of rational possibility includes the effects of non-rational intervention on the agent. Both views on the issue result in a version of Modest Internalism and while I have thus far assumed its inclusion, if there is a compelling argument for the narrower interpretation of 'rational possibility', that argument itself may be added to the justification already given for Modest Internalism in Chapter 5, Section II.

With an analysis of a justifying reason completed, Sections II and III contend that moral facts necessarily entail sufficient justifying reasons to act, P_1 in The Relative Reasons Argument¹¹. Section IV then turns to provide an analysis of particular types of moral facts in a qualifiedly similar form to the analysis of justifying reasons. This analysis also grounds discussion of the degree of objectivity of moral facts in Chapter 8.

I Analysis of a Justifying Reason

If the necessary condition in Modest Internalism is not also a sufficient condition for justifying reasons then something else must obtain or occur for a sufficient condition of justifying reasons. This ‘something else’ could be:

- 1) that someone other than $A+$, namely $B+$, could also recognize that the reason exists;
- 2) that some other fact(s) obtains; or
- 3) that the conjunction of 1) and 2) occurs.

The first two possibilities face the same dilemma and the third is problematic because of the problem facing the first two. On the first possibility, either $B+$ could (or would) judge that the justifying reason exists on a given occasion or she could (or would) not. (Hereafter I omit the addition of the alternative ‘would’ as it does not make a difference to the analysis). If $B+$ *could* judge that the reason exists, then this additional necessary condition is redundant because the initial necessary condition ‘that $A+$ could judge that the reason exists and could be motivated to act in virtue of it’ already suffices to establish that it exists.

So this additional necessary condition will only be relevant where $A+$ judges that she has a reason to Φ in C at t , and $B+$ either dissents or is agnostic on

the matter: that is, $B+$ judges that $A+$ does not have this reason or that she does not know if $A+$ has this reason. However, in these circumstances, if $B+$ is not present and does not communicate her dissent or agnosticism to $A+$ then, $A+$ could not believe that $B+$'s judgment is correct because it is inaccessible to her ($A+$). Alternatively, if $B+$ is present and communicates her dissent or agnosticism to $A+$, her communication forms part of the circumstances (' C ') already stated and, as such, $A+$ would have already taken $B+$'s judgment into account.

Suppose then that the second necessary condition is required. For there to be a reason for A to Φ in C at t , in addition to $A+$ being able to have a certain response to the reason, it is also a necessary condition that some fact or state of affairs obtain. This will not include the state of affairs 'that the reason exists' for it is trivial that a necessary condition for reason R existing is that reason R exists. So we are rather concerned with a necessary condition that some other fact or state of affairs obtains. However, any such fact or state of affairs proposed faces the same dilemma. Either $A+$ could know that the state of affairs exists in virtue of it existing or she could not. If she could know that the state of affairs exists, then the purported necessary condition is already accommodated by $A+$ correctly recognizing it to exist because it exists. There is, of course, a difference between ' $A+$ correctly recognizing that a state of affairs exists because it exists' and 'the state of affairs existing'. However, by definition the former entails the latter, so the difference between the two states of affair does not demand the additional necessary condition.

Alternatively, if $A+$ could *not* know that this particular state of affairs obtains, then the existence of the practical justifying reason is beyond $A+$'s capacity to correctly infer that it exists, and in light of the argument of Chapters 4 and 5, the purported justifying reason could not be a justifying reason at all. Accordingly, there is no further necessary condition for

practical justifying reasons existing, and so, the following biconditional is true:

Justifying Reason: R is a sufficient justifying reason for A to Φ in C at $t \leftrightarrow (A^+$ can recognize R and be motivated to Φ in C at t for R) and (δ)

where ' A^+ ' is a version of A who reasons as well as A possibly could in C at t and where ' δ ' denotes a possible further requirement that some other mental state or response from A is present.¹ With an analysis of justifying reasons in hand, we can consider if there is a necessary relation between moral facts and justifying reasons.

II Morality/ Reason Internalism Disambiguated

To identify the particular variation of Morality/ Reason Internalism we are concerned with, we must firstly note the distinction between moral judgments and facts:

MJ—R Internalism: a sincere *moral judgment* by a rational agent A that either A herself, or another rational agent B , should Φ in circumstances C , at time t , necessarily entails an insufficient pro tanto reason, a sufficient pro tanto reason or a conclusive reason for A or B respectively to Φ in C at t .

MF—R Internalism: a *moral fact* that an agent A should Φ in circumstances C , at time t , necessarily entails an insufficient pro tanto reason, a sufficient pro tanto reason or a conclusive reason for A to Φ in C at t .²

¹ I need not commit to a view on what, if any, additional mental state or response from A is required at this stage so I leave this question aside.

² MF—R Internalism is sometimes called 'existence internalism' in contrast to 'judgment internalism' (what I call MJ—R Internalism), following Stephen Darwall. In order to retain consistently formatted names I adopt the terminology stated here. Darwall, 'Internalism and Agency'.

In the context of The Relative Reasons Argument^{||}, there is an advantage and complementary disadvantage to focusing on each theory over the other. MJ-R Internalism ensures that the agent making the judgment is aware of the consideration that gives rise to the reason to act as giving rise to that reason. Yet, there is no assurance that the agent's judgment about this consideration or about the judgment that is based on it is correct and for this reason there may be no need for Moral Realism to deny it. Conversely, MF-R Internalism ensures that the moral fact purportedly giving rise to a reason to act obtains. However, there is no assurance that a given moral agent could know of the fact in question, and therefore have the reason to act.

Both disadvantages can be accounted for. For MJ-R Internalism, a concern that false moral judgments will entail reasons is accounted for by the distinction between subjective and objective justifying reasons (where a subjective reason is understood in either of the two senses introduced).³ When the agent's sincere moral judgment is correct, the judgment entails a justifying reason that is objective and subjective; when the agent's judgment is incorrect, then, insofar as she sincerely continues to believe it, she has a subjective justifying reason.

In contrast, for MF-R Internalism, an agent's inability to identify certain moral facts is accounted for by confining the doctrine to those moral facts an agent could recognize and act on. Indeed, MJ-R Internalism is already so confined in virtue of merely being concerned about *agents'* moral judgments. This means that, even if The Relative Reasons Argument^{||} is sound, when either version of the theory is in **P₁**, the argument's conclusion cannot deny the possibility of cognitively inaccessible or recognition-transcendent moral

³ Recall from Chapter 1, II the distinction between a subjective justifying reason that exists for an agent relative to the circumstances as she genuinely understands those circumstances to be, and a subjective justifying reason that exists for an agent relative to circumstances that she is justified in believing to exist.

facts, and moral propositions, the truth or falsity of which is agent-neutral. For some academics this worry may be partly assuaged by only caring to realize something about the world as we know it, rather than something that we shall never know about. In any event, Section III in this chapter provides reason to doubt the existence of recognition-transcendent moral facts and for simplicity, hereafter I focus on MF–R Internalism.

One version of MF–R Internalism I set aside is Moral Rationalism (or Metaethical Rationalism),⁴ although some writers reserve this appellation for other theories.⁵ I use it here to denote one or more of the following propositions or a theory maintaining such:⁶

⁴ Smith, *The Moral Problem*, chap. 3; Unger, *Living High and Letting Die*, 21–2; Miller, *An Introduction to Contemporary Metaethics*, 228; van Roojen, ‘Moral Rationalism and Rational Amoralism’; Joyce, *The Myth of Morality*, 80. Notably, however, Joyce adds that Moral Rationalism also claims that the imperatives of practical rationality are categorical imperatives and that there are some true imperatives of practical rationality. Jeffrey Wisdom also uses the term in this way but distinguishes between the claim that moral facts are grounded in facts about practical rationality and the claim that moral failings are rational failings. Wisdom, ‘From Moral Fictionalism to Moral Realism: An Essay in Moral Metaphysics’, 183. Other writers also use the term ‘moral rationalism’ or ‘metaethical rationalism’ in this way, albeit in response to Smith’s arguments: Horgan and Timmons, ‘Troubles for Michael Smith’s Metaethical Rationalism’, 204; Beaulieu, ‘Meta-Ethical Rationalism and the Amoralist Challenge’; Zarpentine, ‘Michael Smith, Rationalism, and the Moral Psychology of Psychopathy’.

⁵ The name ‘moral rationalism’ is often used to denote different but related claims, including: **1)** claims similar to what I call Morality/ Reason Internalism (Parfit, ‘Reasons and Motivation’, 103; Shafer-Landau, *Moral Realism*, 170); **2)** that obligations furnish people with reasons to act which have no special normative force and which have no distinctive role in practical deliberation (‘simple rationalism’) and that ‘it makes sense to do something because you are under an obligation to do it only in so far as this obligation constitutes a reason to do it and/ or a reason for you to deliberate about whether to do it in a certain way’ (‘rationalism about obligation’), (Owens, ‘Rationalism about Obligation’, 404); **3)** that all things considered, acting morally is rationally required (Jussi Suikkanen in Shoemaker, ‘Ethics Discussions at PEA Soup: Mark van Roojen’s “Moral Rationalism and Rational Amoralism,” with Commentary by Russ Shafer-Landau’); and **4)** the epistemic claim that: basic moral principles are known to us a priori (Peacocke, ‘Moral Rationalism’). In addition, Dale Dorsey uses the term ‘moral rationalism’ (and for the converse theory ‘moral anti-rationalism’) to denote either of the first two versions of Moral Rationalism just listed (essentially: that a moral demand entails a pro tanto or conclusive reason for action). Dorsey then defends what he calls Weak

Moral Rationalism:

1. The fact that it is morally right for A to Φ in C is *reducible* to the fact that A has a pro tanto/ conclusive justifying reason to Φ in C ; or
2. The fact that it is morally right for A to Φ in C is reducible to the fact that it is rational/ practically rational for A to Φ in C ; or
3. The demands/ obligations of morality are identical to the demands/ obligations of practical reasons.

Proving the truth of either proposition 1 or 2 also proves MF–R Internalism. Yet, each of propositions 1 and 2 (and 3) is controversial⁷ and P_1 does not require a theory as strong as this: it only requires that a moral fact *entails* a sufficient or conclusive reason to Φ . Thus, when I refer to MF–R Internalism I am concerned with this less demanding version.

Finally, the use of some versions of MF–R Internalism in The Relative Reasons Argument⁸ renders the argument invalid. In particular, The Relative Reasons Argument⁸ may only be valid if the kind of justifying reasons in MF–R Internalism fall within the scope of Modest Internalism (sufficient or conclusive justifying reasons that A + could recognize and be

Anti-rationalism: ‘that a norm n is a moral requirement for a to x provides sufficient reason for a to x ’, although not always decisive reason for a to x . Dorsey, ‘Weak Anti-Rationalism and the Demands of Morality’, 9.

⁶ A corresponding version of this theory can also be identified in regard to MJ–R Internalism.

⁷ For writers defending the claim see: Smith, *The Moral Problem*, chap. 3; Smith, ‘In Defence of The Moral Problem: A Reply to Brink, Copp, and Sayre-McCord’; Quinn, ‘Putting Rationality in Its Place’; Foot, *Natural Goodness*; Skorupski, ‘The Unity and Diversity of Reasons’, 134; Skorupski, *The Domain of Reasons*; Schroeder, *Slaves of the Passions*; van Roojen, ‘Moral Rationalism and Rational Amoralism’. Kant is often thought to be famous for this claim, but see Peter Railton’s comments on this point: Kant, *Groundwork for the Metaphysics of Morals*; Railton, ‘Normative Force and Normative Freedom: Hume and Kant, but not Hume *versus* Kant’, 341–3. For criticism of Moral Rationalism see: Wong, ‘Moral Reasons: Internal and External’, 558; Frankfurt, ‘Rationalism in Ethics’, 259; Frankfurt, *Taking Ourselves Seriously and Getting It Right*; Vogler, *Reasonably Vicious*; Copp, ‘The Normativity of Self-Grounded Reason’, 312; Copp, ‘A Skeptical Challenge to Moral Non-Naturalism and a Defense of Constructivist Naturalism’, 270, 274; Miller, *An Introduction to Contemporary Metaethics*, 228; Unger, *Living High and Letting Die*, 21–2; Milo, *Immorality*, 210–7.

motivated to act on in the relevant circumstances, at the relevant time and for the reason in question).⁸ Thus, it is only prudent to consider the truth of versions of MF–R Internalism concerned with reasons of this kind. In turn, this means that it is unimportant whether moral facts necessarily entail *insufficient* pro tanto justifying reasons ‘from the moral point of view’, or merely in virtue of morality being a normative institution: what are sometimes called ‘institutional reasons’.^{9/ 10} Indeed, it was with awareness of this possibility that we turned to consider **P₅** of The Relative Reasons Argument first.

For this reason I consider whether moral facts necessarily give rise to *sufficient* reasons to act. From Chapter 1, a sufficient reason is a single justifying pro tanto reason to act or a combination of insufficient pro tanto justifying reasons to act which, taken together, are sufficient for one to justifiably act on while not necessarily constituting the optimal course of action. A sufficient reason therefore may, but need not, constitute a conclusive justifying reason.¹¹ As a result, this version of MF–R Internalism differs from the harder line presently and traditionally discussed in the literature: that moral facts entail *conclusive* justifying reasons to act. If true, this latter theory could serve as **P₁** but I have reservations about it and therefore leave it aside.^{12/ 13}

⁸ Brink and Joyce make this point: Brink, ‘A Puzzle About the Rational Authority of Morality’, 8; Joyce, *The Myth of Morality*, 43–5.

⁹ Brink, ‘A Puzzle About the Rational Authority of Morality’; Joyce, *The Myth of Morality*, 44.

¹⁰ There is, I think, a potentially credible argument for thinking that some insufficient pro tanto reasons fall outside the scope of Modest Internalism. See Chapter 6, n3.

¹¹ See Chapter 1, II for the definition of a ‘sufficient justifying reason’.

¹² Some writers defending the view that we always have conclusive reasons to act according to what morality requires include: Plato, *The Republic*; Singer, ‘Famine, Affluence, and Morality’; Singer, *The Life You Can Save*; Kagan, *The Limits of Morality*; Unger, *Living High and Letting Die*.

¹³ In any event, I agree with Mark Eli Kalderon and Jonathan Dancy arguments for believing that non-moral reasons may outweigh moral reasons to act. Kalderon, for

In sum, I defend the following version of MF–R Internalism:

MF–R Internalism: *A (cognitively accessible) moral fact that a rational agent, A, should Φ in circumstances C, at time t , necessarily entails a sufficient justifying reason, R, for A to Φ in C at t where R bears a necessary relation to A's capacity to recognize and be motivated to act for R.*

(However, the moral fact itself need not be *reducible* to the normative fact that 'A has a sufficient reason to Φ in C at t ').

III MF–R Internalism: Justification

Limiting the scope of MF–R Internalism to cognitively accessible moral facts may quite sensibly lead to two concerns. Firstly, without further discussion, this correspondingly limits the scope of this thesis's contention to such facts. In this way, it would fail to address a version of Moral Realism committed to recognition-transcendent moral facts. Secondly, limiting MF–R Internalism to cognitively accessible facts does not entail that the moral facts in question are accessible *by the agent whom the moral fact concerns*. For all that has been said, the limitation of cognitive accessibility is respected if moral agent *B*, but not *A*, could recognize that there is a moral fact that *A* ought to Φ in C at t .

In the face of both worries, it is firstly worthwhile iterating that we are concerned with the substantive question of whether moral facts (and in earlier chapters, moral reasons) exist and in precisely what form they exist. We are not concerned with the question of what pre-reflective conception of these entities we have.

example, argues that a light promise that is not of much consequence may give way when a rare and important opportunity arises. Kalderon, *Moral Fictionalism*, 27–8; Dancy, *Ethics Without Principles*, 43.

Secondly, both concerns are addressed by doubts we should have about the foundation from which they start to gain purchase because both worries are based on an assumption that moral *facts* exist in all instances in which common sense morality might suppose that they do. However, for reasons analogous to those discussed in Chapter 6, this assumption is false.

Thus, one simple argument for there being a moral fact, *M*, about a rational agent, *A*+, who cannot recognize *M* in *C* at *t* after reasoning as well as *A* possibly can, may take the following form:

A moral fact about *A*'s Φ -ing in *C* at *t* may exist without there being a sufficient reason for *A* to Φ *C* at *t* that *A* can recognize, just because it is true that *A*'s Φ -ing *C* at *t* is morally right.

This argument should get our memory stirring. We rejected this argument in Chapter 6 in regard to justifying reasons and the same justification applies here: whether or not it is *true* that *A*'s Φ -ing in *C* at *t* is morally right is the issue in dispute.

A second objection that could be put to MF–R Internalism is as follows:

How can it be true that a moral fact necessarily entails a moral reason to act for that reason, in cases where:

- a) the moral fact gives agent *A* a sufficient justifying reason *R* to Φ in *C*, at *t*, *solely* for *R*; and
- b) the content of *R* does not include moral terms?

In such circumstances, acting solely for *R* is not acting under the heading of *R* being a *moral* reason.

The type of case in question is well known and simple to illustrate. Marcus's wife ingests a wild mushroom and unsurprisingly starts feeling violently ill. As a result, suppose there is a moral fact that it is morally right for Marcus to save his wife's life and that he has a sufficient (and conclusive) justifying

reason to do so. Nevertheless, something is amiss if Marcus saves his wife's life either for the reason that: **i)** it is morally right to try and save her; or for the reason that **ii)** he loves her *and*, because he loves her, it is morally right to try and save her.¹⁴ So, the objection maintains, in this case it is morally right that he try and save her, but there is no justifying reason for him to save her *for the reason that this is morally right*.

This objection is analogous to the circumstances in which Archie may have a reason to 'stay away from the funeral because there is an interim restraining order against him' and indeed he may have a reason to 'stay away from the funeral because he may make things worse'. However, at least so far as I maintained, Archie cannot have a reason 'to stay away from the funeral because he may make things worse *because* he is insensitive'. Similarly, Nate may have a reason to 'enter the living room to catch up on *Sky & Telescope*'. In exceptional cases, he may even have a reason to 'enter the living room because there is a surprise party lying in wait': there is, for example, something about this party that outweighs his dislike for unsuccessful attempts at surprising him. However, at least so far as I argued, he cannot have a reason to 'enter the living room because there is a surprise party lying in wait *and* because he wants to be (entirely) surprised by it'. In the case of the poisonous mushroom, Marcus may have a reason 'to try and save his wife's life because he wants to learn how to do so' and he may have a reason 'to try and save his wife's life solely because he loves her'. However, he cannot have a reason 'to try and save his wife's life solely because he loves her *and* because it is morally right to do so'.

Thus, in light of our discussion of Archie and Nate we can see that there is no problem with cases such as Marcus'. One of two answers is available. On one answer, we may agree that there is a moral fact that it is morally right for

¹⁴ A third possibility is available but does not add anything new to the discussion: **c)** because he loves her and because it is morally right to try and save her: Williams, 'Persons, Character and Morality', 18.

Marcus to save his wife solely because he loves her. We may also agree that there is no justifying reason for Marcus 'to save his wife solely because he loves her and because it is morally right' because neither Marcus nor anyone else could recognize and act on this reason. Yet we may deny that MF-R Internalism requires that moral facts necessarily entail *moral* justifying reasons for relevant agents to act. We may instead claim that moral facts only necessarily entail either moral or non-moral justifying reasons to act, in this case, a non-moral reason for Marcus 'to save his wife solely because he loves her'.

Alternatively, we may agree that there is a moral fact that it is morally right for Marcus to save his wife yet contend that the justifying reason for Marcus 'to save his wife solely because he loves her' that follows, *is* a moral reason, it is just that it need not be comprehended and acted on under the guise of being a moral reason. The *moral* reason for Marcus 'to save his wife solely because he loves her' may be one and the same as the non-moral reason for Marcus 'to save his wife solely because he loves her'. It is just that the former reason is understood in a different light or from a different perspective. In this way, it is possible for Marcus to recognize and be motivated to act on the *moral* reason 'to save his wife's life solely because he loves her'.¹⁵ Hereafter I follow this latter answer for ease of expression: this answer allows me to use the term 'moral reason' both in regard to cases where moral considerations enter the content of the justifying reason and cases where they cannot.

The final objection should also be evocative:

A moral fact may exist that does *not* necessarily entail that one has a sufficient reason to Φ in C at t , because even if A does not have a reason to Φ in C at t , it is still plausible to maintain, and we would still maintain, that A is responsible for not Φ -ing in C at t where, for

¹⁵ Compare Brian Loar's reply to The Knowledge Argument: Chapter 2, V.

example, Φ -ing involves little effort and cost to A and would, for example, save the residents of a small town.

Unsurprisingly, a similar response given to this argument put to moral reasons arises can be given here. Thus, if in the hypothetical given, at the present point in time (call this, t_0) A does not have a reason to Φ in C_1 at some later point in time (say, t_5) and A shall not have a reason to Φ in C_1 at t_5 from $t_0 - t_5$, then, firstly, from Modest Internalism we know that, from $t_0 - t_5$ A cannot recognize or be motivated to Φ in C_1 at t_5 . We may nevertheless hold A responsible for not Φ -ing in C_1 at t_5 (for not, for example, saving the lives of those in a small town) in virtue of her decisions and actions prior to t_0 that led her to be in her present state in which she cannot recognize or be motivated to save those in the town after reasoning as well as she could. Whether we can hold her so responsible or not depends on whether those prior decisions and actions were indeed ones for which she could have recognized a reason to avoid, if she reasoned as well as she could. If not, and if we cannot refer to some even earlier point in time in which this is true, then no, we cannot hold her responsible. Furthermore, this result makes sense in virtue of the stipulation that A could never have reached a state from which she could recognize a reason to Φ in C_1 at t_5 when reasoning as well as she possibly could. There will, of course, be an explanation for A 's abnormal development and it may then be possible to hold some one or more persons responsible for her benightedness (for instance, Murder Inc. in Gilbert Harman's illustration).¹⁶

In contrast, if her earlier decisions and actions were ones for which she could have recognized a reason to avoid and been so motivated, then we may hold her responsible for the consequences of this, including her not Φ -ing in C_1 at t_5 (i.e. her not saving the townsfolk). Notice, however, that when the hypothetical is described in this way, it does not amount to a

¹⁶ Harman, 'Moral Relativism Defended', 5.

counterexample. Rather, depending on further details, either: **1)** the moral fact that ‘it is morally wrong for *A* to omit Φ -ing in C_1 at t_5 ’ does not exist, but the moral fact that ‘it is morally wrong for *A* to ψ or omit ψ -ing in C_2 at t_{-15} or t_{-10} ’ exists and entails that there is a sufficient justifying reason for *A* to ψ in C_2 at t_{-15} or t_{-10} ; or alternatively: **2)** the moral fact that ‘it is morally wrong for *A* to omit Φ -ing in C_1 at t_5 ’ exists and entails a moral reason for *A* to act in virtue of this. It is just that this moral reason is not one to Φ in C_1 at t_5 , but one to ψ or omit ψ -ing in C_2 at t_{-15} or t_{-10} *so that* she could (or so that it might be that she could) see a moral reason to Φ in C_1 at t_5 . The clearest illustration of this second case may be cases of wilful blindness. For instance, *A* knows that she has an obligation to perform a certain act at or by t_5 (blow the whistle on her organization; give someone bad news; throw a fight in exchange for saving someone’s life; appear to betray a friend), but to avoid doing so she takes a sleeping pill at t_{-10} and is not conscious from t_0 - t_5 .

Accordingly, for the same reasons that we may doubt the existence of external sufficient moral reasons, we may doubt the existence of moral facts that do not at least entail sufficient moral reasons to act. The result is that moral facts and moral reasons are necessarily coextensive in the actual world. This in itself proves that MF–R Internalism is true because it establishes that a biconditional equation obtains between moral facts and moral reasons. However, while we have seen *why* the two are in fact co-extensive, this does not explain why *an agent* may think that she has a sufficient reason to Φ because a moral fact obtains (or indeed, *vice versa*). Without explaining this, we may be left feeling like an important piece of the puzzle is missing.

Thus, when it is morally right for *A* to Φ , why is it that *A*+ can recognize and be motivated by this moral fact and a sufficient reason *R* to Φ , where the content of *R* is that it is morally right to Φ ? Significantly, the answer need not consist in one evaluative attitude, interest, reason etc. common to all moral

agents in circumstances to which our discussion is limited. As mentioned in discussing The Humean Theory of Motivation, all that is required is that, necessarily, in each case one such resource or another *does* obtain.¹⁷

Some such sufficient moral reasons may arise from the agent's 'self-interest' and, since we are only concerned with cases where the agent is reasoning as well as she possibly can, I think such a resource, properly defined, provides a more extensive explanation for the co-extensiveness of moral facts and sufficient moral reasons than is usually granted. However, firstly, Richard Joyce may be right in observing that self-interest nevertheless fails to explain the existence of, in this context, subjective sufficient moral justifying reasons in near and far possible worlds.¹⁸ Secondly, as H. A. Prichard argues in a different context, self-interest is not always the right *kind* of explanation for why moral agents are able to recognize moral reasons to act.¹⁹ Thus, even if self-interest could explain how moral agents can recognize and be motivated by moral reasons where they can in all cases in the actual world (and hence, provide the missing piece of the puzzle), a better explanation is available.

In cases where *A*+ can recognize and be motivated to Φ for a sufficient moral reason, the simpler explanation is that she can do so, firstly because she can appreciate the reality of the first and third-person points of view of other moral subjects, as well as an impartial or 'moral' point of view, that does not derive from the interests of any particular sentient creature. The impartial point of view is one that values all moral subjects and their points of view equivalently, where this is appropriate (even if this impartial point of view is only conceivable from a given agent's personal point of view).

¹⁷ H. A. Prichard and Gavin Lawrence also make this point in regard to MJ-Reason in virtue of personal gain. Prichard, 'Duty and Interest', 212; Lawrence, 'The Rationality of Morality', 99.

¹⁸ Joyce, *The Myth of Morality*, 60.

¹⁹ Prichard, 'Does Moral Philosophy Rest on a Mistake?', 5–13.

This, in itself, does not suffice to explain one's going beyond acting out of self-interest, where by 'self-interest' I mean not acting for a moral reason. A given agent may recognize the point of view of other moral subjects, or the impartial point of view, as well as what such views entail, and yet still act from self-interest because it is her interest.²⁰ What makes it the case that *A+* can recognize a moral fact that subsequently gives her a sufficient moral reason to Φ in *C* at *t* because it is a moral reason, is her also believing that, in *C* at *t* another moral subject's point of view actually matters, or acting from the impartial point of view actually matters.

In this respect, the charge that *A+* is better off acting in her own interests (and where relevant the interests of those she favours), instead of acting from a more impartial point of view is irrelevant. *A+*'s acting out of self-interest does not mean that she cannot have a *sufficient* moral reason to perform a morally right act. In addition, where *A+* could not recognize or be motivated by a sufficient moral reason to Φ in *C* at *t* then, as we have seen, relative to *C* and *t*, no moral fact about her Φ -ing for this reason obtains, and no such moral reason obtains. Hence, the charge misunderstands the context of the discussion. We are only concerned with circumstances in which *A+* *could already recognize and be motivated by a certain moral reason*. This occurs (when it does occur), simply because *A+* recognizes the reality and importance of other agents' points of view and a more impartial point of view. Indeed, this may be, in itself, what explains her *capacity* to recognize and be motivated to act for moral reasons because a moral fact of the matter obtains.

²⁰ This is Sergio Tenenbaum's contention: Tenenbaum, 'Ethical Internalism and Glaucon's Question'.

IV Moral Facts & Recognition

A similar biconditional to Justifying Reason does not obtain between *moral facts* and agents. We have seen that if a moral fact exists, it necessarily entails a sufficient justifying reason for *A* to perform or not perform a certain act. In virtue of the truth of Justifying Reason, we also know that *A*⁺ could recognize and be motivated to act for that justifying reason, where ‘for’ means that *A*⁺ knows that she is acting in accordance with the moral fact and reason in question, or at least what she perceives to be a moral fact and reason. For all this, however, in some circumstances there may be no fact of the matter as to whether the act called for is a ‘moral’ (as opposed to non-moral) act. This may be because the matter is vague or the distinction is false. I leave this for normative ethics to answer as well as the question of what fact(s) determines that a matter is moral or non-moral. The point is that since these are live issues, any given rational agent, *A*⁺, cannot necessarily know the answers herself and hence, *A*⁺ may (falsely) believe that in circumstances *C* at *t*, there is a moral fact about her Φ -ing, when actually no such fact exists. In this case, *A* will have a sufficient subjective justifying reason to Φ in *C* at *t*, because, as we have discussed, *A*⁺ may not know that the matter is a non-moral one. However, her mistaken belief that the moral fact exists does not carry so much weight as to make it the case that the matter *is* a moral one. Thus, suppose that on the basis of certain information, *A*⁺ believes that, morally, she should smile whenever she passes someone on the street in Giverny, France. In virtue of her belief, *A* has a sufficient justifying reason to do so. Yet for all this, since her belief is false, there is no moral fact about her conduct. In light of the possibility of cases of this kind, and other kinds, moral facts are not coextensive with *A*⁺’s sincere *determination* that a moral fact exists.

We can account for this by providing in the biconditional that ‘ Φ -ing *is a morally relevant act*’ and by discussing specific moral values, such as the

properties of 'being morally right' or 'being morally wrong' (compared to the genus 'moral fact'):

Morally Right: It is morally right for A to Φ in C at $t \leftrightarrow \Phi$ -ing is a morally relevant act & (A + could recognize and be motivated to Φ in C at t for sufficient justifying reason R) & (δ)

Morally Wrong: It is morally wrong for A to Φ in C at $t \leftrightarrow \Phi$ -ing is a morally relevant act & (A + can recognize and be motivated to avoid Φ -ing in C at t for sufficient justifying reason R) & (δ)

As in Justifying Reason, in each of these two biconditionals, 'R' denotes a sufficient justifying reason for A to Φ in C at t that is entailed by the moral fact. Similarly, ' δ ' denotes a possible further requirement that some other mental state or response from A is present.

Each biconditional is plausible. Firstly, neither biconditional is rendered trivial by the proviso that ' Φ -ing is a morally relevant act' since this fact is neutral on the question of whether Φ -ing is *right* or *wrong* (or good or evil etc.).

Secondly, the capacity to determine that Φ -ing is or is not a morally relevant act also does not illegitimately presuppose that a *property* of some kind (perhaps even and most concernedly, a moral property) is required to 'make' this determination true. The determination that Φ -ing is a morally relevant act could consistently be established by A + making a *further* judgment about this matter or alternatively by a metaethical fact.

Thirdly, if the determination that ' Φ -ing is a morally relevant act' is itself a moral fact, then the biconditional will be non-analytical: it will not successfully analyse a moral concept in non-moral terms. It is often pointed out that this in itself is not a problem if the aim of the analysis is to elucidate the relations between concepts, rather than to outline conditions for the

analysis of one concept in different terms.²¹ The aim of elucidating relations between concepts is still a means to an end, and more than one possible end is available. So, more specifically, the aim of elucidating relations between concepts in the two biconditionals here is to identify the truth conditions of moral evaluations.

Finally, where *A*+ must choose between 'the lesser of two evils', each act might be 'morally wrong' on the analysis given and, insofar as we consider that at least one act *must* be right, it might be proposed that this result challenges Morally Wrong. However, the result that each option is morally wrong is context-dependent. Insofar as each act is 'evil', then it might very well be that each is morally wrong. However, insofar as one act must be right, despite its evilness, then there will also be a sufficient justifying reason to perform that act, in which case it will be morally right. The same answer applies, *mutatis mutandi*, to circumstances where it is considered problematic that each of only two alternative acts is morally *right*.

Nonetheless, even if correct, in virtue of the analyses in each biconditional including the possibility of a certain response by *A* (in this case, two responses in each biconditional), Morally Right and Morally Wrong may be regarded as falling within a class of theories in philosophy sometimes called 'response-dependent' theories. As with all theories of this kind, it remains to identify whether the left or right side of the biconditional has explanatory priority or whether explanatory priority is interdependent or nonexistent (co-variation in truth may be ensured by some third fact or God). We turn to this issue next.

²¹ Wiggins, 'A Sensible Subjectivism?', 187–9; Johnston, 'Dispositional Theories of Value', 147–8; Pettit, 'Realism and Response-Dependence', 604.

8

Moral Values & Agents

The big thing isn't big because we think it's big. It *is* big.

—Jason Ensler, *The West Wing*, 2.20

Chapter 1 established that one of the two main points dividing writers using the title moral realism is whether the theory is committed to moral facts or moral propositions existing independently, in some sense, of moral agents. A number of writers use the name in this way, and this more robust version of moral realism is also defended by several philosophers.¹ However, if MF-R Internalism and Modest Internalism are true, as I have argued, then moral facts cannot be independent of rational human agent's cognitive and motivational capacities. In virtue of this it is a condition of the existence of moral facts that rational human agents exist, and have the motivational capacities that they do, even if it is not the case that this relation of dependence is agent-relative. As a result, when moral realism is defined robustly in this way, the theory is false.

¹ Shafer-Landau, *Moral Realism*; Oddie, *Value, Reality, and Desire*; Huemer, *Ethical Intuitionism*; FitzPatrick, 'Robust Ethical Realism, Non-naturalism, and Normativity'; Enoch, 'An Outline for an Argument for Robust Metanormative Realism'.

There is however space for versions of moral realism that allow and defend the idea that moral facts (in some sense of 'fact') are dependent on the existence of rational human agents. Two types of theory of this kind are defended in the literature in regard to both metaethics and philosophy generally, each of which is sometimes referred to as 'response-dependent' theories. In what follows I consider these two type and contend that theories in the first type which give greater explanatory weight to moral facts lack sufficient justification, while a theory from the second type that I call Judicious Agent Theory is at least plausible.

I Response-Dependent Theories: Extension-Reflecting & Extension-Determining Accounts

Response-dependent theories may take different objects: correct judgments falling within a certain wide class; possession conditions or, alternatively, correct application conditions for concepts; and properties.² The left-hand side expression in Morally Right and Morally Wrong³ may be interpreted as either the concept of 'being morally right/ morally wrong for A [etc.]' or the property of this. For ease of expression I proceed by discussing such equations in terms of properties. I use the term 'property' here in the wide sense that Frank Jackson uses to denote simply a pattern that allows projection.⁴ In addition, I take this to include relations as well as 'mind-

² For example, Philip Pettit argues that, for all concepts, the conditions for possessing a concept are response-dependent. He distinguishes this view from the view that the correct application of a concept requires a certain response, maintaining that one may possess a concept without consciously knowing the conditions for its correct application: Pettit, 'Realism and Response-Dependence', 595–606. See also: Jackson and Pettit, 'Response-dependence without Tears', 97–105.

³ See Chapter 7, IV.

⁴ Jackson, 'What Are Cognitivists Doing When They Do Normative Ethics?', 95.

dependent' entities.⁵ This approach may leave open the possibility that either the possession or correct application of a concept N is response-dependent while, in a certain sense, the property N that concept N stands for is not. This possibility emerges if one acknowledges that property N, for instance, may be contingently identical to property P. Property N may then not be response-dependent, insofar as property P is not response-dependent, while *concept* N is. While I doubt the truth of theories countenancing contingent property identities (anti-conceptualist theories of properties), I need not discuss the matter here: I mostly focus on properties that concepts patently stand for and where such a property may be identified with another, I simply acknowledge this and thereby draw attention to the possibility of a discrepancy arising between the concept's status as response-dependent and the property's status in this regard.⁶

The basic response-dependent schema-template with which I am concerned then is as follows:

RD Property: It is a priori true that:

An entity E or an act Φ -ing bears property N \leftrightarrow A type of agent A has response R_P in suitable conditions C at time t in regard to E or Φ -ing.

where R_P , A, C and t are not specified in a way that renders the biconditional trivial. Even within metaethics, response-dependent theories vary in a number of ways according to how the variables in RD Property or similar

⁵ In understanding a 'property' as 'a pattern that allows projection', I remain uncommitted to a stance on the issues of: **a)** whether properties are universals or tropes or both at the same time; **b)** whether properties are categorical or dispositional, whether some are categorical and others dispositional or whether they are both at the same time; and **c)** whether some properties are structured.

⁶ For a brief discussion of the issue see: Blackburn, 'Circles, Finks, Smells and Biconditionals', 262–3.

schema-templates for response-dependent *concepts* are described.⁷ Nevertheless, as a start we can say that the applicability of RD Property is illustrated by properties that are transparently response-dependent, such as: being seductive, exasperating, chic, rank, loud, amusing, relaxing, spicy, and spine-tingling. We think that the existence of each of these properties is (modally and essentially) *dependent* on certain responses from human beings such as the sensation of spiciness. But we also consider that we only experience this sensation in virtue of *an object*—the Jamaican hot pepper that is spicy; the caricature that is humorous—and often a certain physical *property* of that object—the capsaicin and dihydrocapsaicin molecules in the pepper, for example. Since the object and physical properties of it are (modally and essentially) *independent* of human beings' existence, response-dependent properties possess both objective and non-objective dimensions.⁸

⁷ Thus, response-dependent accounts vary in regard to the subject of the account insofar as some writers, such as Phillip Pettit and Frank Jackson, claim that the account can be given to all concepts. Pettit, 'Realism and Response-Dependence'; Jackson and Pettit, 'Response-dependence without Tears'. For criticism see: Smith and Stoljar, 'Global Response-Dependence and Noumenal Realism'. The subject of the analysis also varies insofar as some accounts are accounts of a given entity E with concept F ('E is F'), in contrast to the account in the above schema-template ('F applies to E') and accounts wherein the subject is a property or relation rather than a concept or general term. Jackson and Pettit, 'Response-dependence without Tears'. Accounts also vary on the whole according to whether the 'account' does not admit circularity or does admit circularity or partial circularity in light of the account's aim of, for example, explicating the concept's relations to other concepts. Typically response-dependent accounts are non-reductive. For example: Smith, 'Dispositional Theories of Value'; Johnston, 'Dispositional Theories of Value'; Pettit, 'Realism and Response-Dependence'. Alternatively, response-dependent accounts of a concept or property are sometimes also defended as empirical truths. For example: Brower, 'Dispositional Ethical Realism'. Some accounts also add further side-conditions to the schema they provide. For example, Crispin Wright adds that the conditions are logically independent of the truth of the application of the concept in question and that the best way of explaining the co-variation of the truth of the concepts is by their dependence on the agent's judgment. Wright, *Truth and Objectivity*, 120–4.

⁸ Thus, for example, Railton explains that values are objective in the sense of this first dimension: 'Although relational, the relevant facts about humans and their world are objective in the same sense that such non-relational entities as stones are: they do not depend for their existence or nature merely upon our conception of them.' Railton, 'Moral Realism', 183.

What is left in dispute is an explanation of why the truth values of the two expressions in RD Property covary. Adapting Crispin Wright's terminology, I describe a theory giving explanatory priority to the left-hand side expression in RD Property ('An entity E or an act Φ -ing bears property N ') as *extension-reflecting*.⁹ Thus, on this explanation it is the existence of the property N that explains why A has R_P in C at t in regard to E or Φ -ing.

In contrast, I describe a theory awarding explanatory priority to the right-hand side expression in RD Property as *extension-determining*.¹⁰ On this account, it is A 's response or judgment in C at t in regard to E or Φ -ing that makes it the case ('determines') that E or Φ -ing bears property N : the big thing is big, because we think it's big, but it is still big for all that.

Wright was also aware of the possibility of the two expressions being explanatory interdependent. I also describe occurrences of this kind as 'extension-reflecting' given the acknowledgment that property N plays a role in making it the case that A has R_P in C at t . In what follows I consider these two types of explanation in turn: extension-reflecting and extension-determining. We shall see that extension-reflecting theories face difficulties demonstrating that moral properties possess the explanatory relevance the theory attributes to them and that as a result an extension-determining theory is more plausible.

⁹ Wright, *Truth and Objectivity*, 110. The initial biconditional that Wright works with differs from RD Property. It is stated in terms of judgments, not properties:

The Basic Equation:

For all S, P : P iff (if CS then RS).

where S is any agent, P ranges over all of some wide class of judgments, RS expresses S 's having of some germane response, and CS expresses the satisfaction of certain conditions of optimality on that particular response. Ibid., 108–9 (bold emphasis present).

¹⁰ Wright, *Truth and Objectivity*, 110.

A. Sensibility Theory

Accounts of RD Property that either give priority of explanation to the left-hand side expression or neither expression are defended by John McDowell and David Wiggins with respect to values generally:¹¹

Values are not brutally there — not there independently of our sensibility — any more than colours are: though, as with colours, this does not stop us supposing that they are there independently of any particular apparent experience of them.¹²

Similarly, after endorsing a biconditional in something of the form of RD Property in regard to values,¹³ David Wiggins writes:

They [value responses] are responses that are correct when and only when they are occasioned by what has the corresponding property φ and are occasioned by it because it *is* φ .¹⁴

As we can see, Wiggins denies that values exist relative to an agent's belief (or agents' beliefs) that an entity x is good/ bad.¹⁵ Furthermore, he denies that values exist relative to a given moral society or relative to a given system of moral assessment.¹⁶ Rather, he maintains, an object x only really has a certain value property φ if it is such as to evoke and make appropriate the response R_P among those who are sensitive to φ -ness, and of course, only when x

¹¹ Strictly speaking, I would say that such accounts are *arguably* defended by these two writers because in these earlier works Wiggins and McDowell do not (expressly or implicitly) distinguish between a schema-template consisting of the left-hand side expression in RD Property ('An entity E or an act Φ -ing bears property N ') and an alternative schema-template consisting in the left-hand side expression: 'Concept N correctly applies to entity E or an act Φ -ing (and hence E or Φ -ing bears property N)'. Nevertheless, their arguments may be understood in light of RD Property.

¹² McDowell, 'Values and Secondary Qualities', 120.

¹³ Wiggins, 'A Sensible Subjectivism?', 195.

¹⁴ Ibid., 204–5 (emphasis present). See also: Ibid. 202. Wiggins also endorses this view in his earlier work: Wiggins, 'Truth, Invention and the Meaning of Life', 106, 132.

¹⁵ Wiggins, 'A Sensible Subjectivism?', 204–5.

¹⁶ Wiggins, 'Moral Cognitivism, Moral Relativism and Motivating Moral Beliefs', 72–7; Wiggins, 'Objectivity in Ethics; Two Difficulties, Two Responses', 17–26.

really has the property of ϕ -ness. So the sense in which a value is objective according to Wiggins, is not just the sense in which explaining one's reaction to something with that value requires referring to the value itself, the value also exists objectively in the sense that it only exists if the response it evokes is 'appropriate' or 'owing' and not merely appropriate for the given agent herself, or her society or a particular system of moral assessment, but appropriate *tout court*.¹⁷ McDowell's view is similar, although he is more willing to call his view 'realism' or at least 'anti-anti-realism'.¹⁸

We can identify two arguments within Wiggins' work in favour of an extension-reflecting theory. Firstly, Wiggins argues that the pretension of objectivity in our use of moral language outside of philosophy places the burden of proof on writers denying the existence of moral properties.¹⁹ There is disagreement about whether or not moral language has this purport.²⁰ Nevertheless, linguistic evidence only has weak probative value and even if this evidence is thought to favour moral objectivity, this does not mean that it only favours Wiggins' Sensibility Theory: the purported objectivity of moral values is also readily explained by certain extension-determining theories, such as certain ideal observer theories and cognitive irrationalism.²¹ For example, one could argue that my judgment that 'slipping an opiate into people's drinks for entertainment is morally wrong' is true in virtue of the fact that our ideal selves would judge it wrong or in virtue of what it is rational not to do. Arguably, on these two explanations, there is a moral 'fact' of the matter (that is more or less metaphysically robust according to different writers) that applies to all moral agents irrespective of their

¹⁷ Wiggins, 'A Sensible Subjectivism?', 203.

¹⁸ McDowell, *Mind, Value, and Reality*, viii.

¹⁹ Wiggins, 'A Sensible Subjectivism?', 201.

²⁰ See, for example: Finlay, 'The Error in the Error Theory'; Joyce, 'The Error in The Error in the Error Theory'; Finlay, 'Errors upon Errors'.

²¹ Of course, irrationalist theories also give explanations for the purported objectivity of moral values.

personal inclinations. Yet, on these two theories, the relevant moral fact is not the ultimate source that determines or partly determines that the act is wrong. Accordingly, Wiggins' assertion that extension-determining theories bear an onus of proof is unpersuasive.

Secondly, Wiggins identifies certain problems for ideal observer theories with a view to motivating acceptance of the need to suppose the existence of moral facts with some explanatory weight. I consider these objections below.

Within McDowell's work we can identify three arguments purporting to establish that moral judgments are extension-reflecting.²² The first makes the negative but important point that a lack of causal efficacy of values cannot be a relevant feature of one's concern for the explanatory adequacy of value properties. Colour and value properties differ insofar as colour properties elicit the appropriate response whereas value properties *merit* certain responses.²³ McDowell writes that in explaining our own value judgments we are also seeking 'to make sense of', so far as possible, what is explained (or justify what is explained) and hence insofar as our explanations have this purport, we must allow that they are also susceptible to criticism.²⁴ If we accept this point we may be satisfied that if value properties do not generate causal effects other than, in a certain sense, value responses, this is not a mark against their genuine existence in the world but merely an implication of their inherent normativity.

McDowell's second and third arguments purport to arrive at a more positive conclusion, yet they are less persuasive. Firstly, McDowell infers from the supposition just described—that in explaining our responses to values we are also seeking to make sense of those responses, so far as possible—that:

²² McDowell, 'Values and Secondary Qualities', secs. 4, 5.

²³ *Ibid.*, 118.

²⁴ *Ibid.*, 119. Wiggins also asserts this point. Wiggins, 'A Sensible Subjectivism?', 199–200.

In so far as we succeed in achieving the sort of understanding of our responses that is in question, we do so on the basis of preparedness to attribute, to at least some possible objects of the responses, properties that would validate the responses.²⁵

Unfortunately, this is all McDowell says on this point: in the remainder of the section he addresses related, but different, concerns.²⁶ Yet, it is not obvious that even if in ordinary discourse we are prepared to attribute properties that would validate the response we have, that this means that our responses are dependent or partly dependent on the existence of those properties. More significantly, even if this were true of ordinary discourse it would not establish that such properties exist after metaethical reflection.

McDowell's third argument in favour of extension-reflecting value judgments emerges in his response to Simon Blackburn's projectivism. McDowell argues that because we cannot step back from the processing mechanism that we inherently employ to try to form correct value judgments and responses, all that is left to prove that one has correctly formed such a judgment is a set of principles or a value property itself. Hence, McDowell's concern is that without such a definite metaphysical basis for validating that one's forming a value judgment is done correctly, we lose the ability to do so at all. Yet notably, McDowell essentially answers a version of this concern himself earlier in the same paper. After explaining that a moral agent may be in error in believing that a certain act is morally right (in virtue of perceiving a genuine moral value in the world), McDowell responds to a concern for one's inability to *ever* be sure that one is making the correct judgment:

²⁵ McDowell, 'Values and Secondary Qualities', 119.

²⁶ In his subsequent comments in this section McDowell explains: **1)** that our explanation for why one evaluation is warranted over another in a given case is itself given from the evaluative point of view; **2)** that the point just described (in **1)** does not entail bootstrapping; and **3)** that this point in **1)** does not mean we cannot have confidence in the explanations we give. *Ibid.*, 119–120.

But although a sensible person will never be confident that his evaluative outlook is incapable of improvement, that need not stop him supposing, of some of his evaluative responses, that their objects really do merit them. He will be able to back up this supposition with explanations that show how responses are well-placed.²⁷

Thus, while McDowell grants moral properties partial or full metaphysical explanatory priority, ultimately he grants epistemic explanatory priority to the agent's reasons for supposing that a certain act is morally right. As McDowell explains here, this means that one can be confident in one's value beliefs. Yet, if this is so, then it is not *necessarily* the case, as McDowell suggests, that the genuine existence of value properties in the world is required to prove that one has correctly formed a value judgment. It may be that an extension-determining theory can account for the truth of true value judgments and as Blackburn observes, such other theories must first be tried out to see if they can provide the requisite explanation before being written off.²⁸

B. Indications of Substantive Moral Properties from Dimensions of Objectivity

While McDowell and Wiggins' arguments fail to prove that moral judgments are extension-reflecting, the problem each argument runs into does not go so far as to establish that extension-reflecting theories are false. However a consideration of the various senses in which moral properties may be relatively objective or non-objective in conjunction with the difficulty of justifying the need for an extension-reflecting theory, prescribes extension-reflecting theories with the onus of proof on this issue.

In defining RD Property we have seen that response-dependent properties possess both non-objective and objective dimensions: response-dependent

²⁷ Ibid., 120.

²⁸ Blackburn, 'Errors and the Phenomenology of Value', 18.

properties are non-objective insofar as it is a condition of the existence of the property that there is or could be a certain response from a subject, such as the sensation of seeing something as the colour violet, the sensation of spiciness, of relaxing. They are also objective insofar as the subject only experiences this response in reaction to an object and often a certain physical property of that object where each entity is (modally or essentially) independent of the existence of the species to which the subject belongs.

In addition to this, response-dependent properties may be relatively objective and non-objective in other senses and this is unsurprising given that 'objectivity' is a term of art. Thus, firstly, a response-dependent property N may be objective in the sense that it would continue to exist if humanity's present responses to the entities bearing it were to change significantly. This may be because, for example, the property is identical to **i)** a macro or microphysical property in the actual world, **ii)** a disposition in a property or object in the actual world over and above the object itself and its physical properties, or **iii)** a certain human response in the actual world now (even if that response changes in the hypothetical world in question).²⁹ In a slightly similar vein, Michael Smith uses the term 'objective' to describe moral requirements as being expressive of reasons for action that are binding on rational creatures.³⁰

Thirdly, moral properties may be considered even more objective/ slightly less non-objective than these first two possibilities in the sense that the property would exist in a world even where it is stipulated that no human (or creature with certain similar characteristics) would be present to respond

²⁹ Richard Joyce, for example, distinguishes between the existential mind-independence of a property and its conceptual mind-dependence, where the former is described as the fact that the property is a disposition in a certain object. Joyce, 'Moral Anti-Realism', sec. 5; Joyce, 'The Accidental Error Theorist', 158.

³⁰ Smith, 'Objectivity and Moral Realism: On the Significance of the Phenomenology of Moral Experience', 250.

to it in the relevant way (i.e. the property is modally independent of humanity's existence). (The property may still be response-dependent here if it is a condition of the property's existence that: if, *per impossibile*, a certain creature were to observe it in the right conditions, it would have the appropriate response).

Fourthly, a response-dependent property may be objective insofar as we are less willing to allow that two parties disagreeing about the status of a particular sensation may both be right. For example, loudness is relatively objective in this sense because if one person maintains that a train whistle (~90 decibels) is soft, then ordinarily we do not think she is right, but that she is being ironic, disingenuous, or that she has a problem hearing.

Finally, Wright uses the term objective to describe an entity with a 'wide cosmological role'.³¹ According to Wright, a property has a wide cosmological role if it can contribute to at least some kinds of explanations of things other than via the specific human response in consideration. For example, the hardness of the bench may not only explain the sensation of its resistance on my hand, but also its capacity to support a large statue. In contrast, the property of being spine-tingling may explain little if anything other than a particular sensation I might feel when reading a murder mystery novel while guarding the Hermitage Friday night, for example.

In addition to these five further senses of objectivity, there are an infinite number of variations of the first further sense that may arise because a property could be objective insofar as it would continue to exist in the actual world even if object, action or event *x* changed, where '*x*' may represent any object, action or event whatsoever (an ant rests for half of a second at *t* compared to not doing so; I touch the window with my index finger at *t*

³¹ Wright, *Truth and Objectivity*, 196–8.

exerting pressure p_1 instead of pressure p_2 ; I lock the door compared to leaving it closed and unlocked). Of course, the occurrence of one of these events compared to its complement may or may not (subtly or significantly) influence the occurrence of subsequent events. The point is not that this is not true; the point is that regardless of how the circumstances are spelt out, we may say that the property of moral wrongness identified in the judgment ‘murdering an innocent person for curiosity in C at t is morally wrong’ is objective in the sense that if one of the above events occurred rather than its complement, the value of the moral property identified in this judgment would not change *solely* in virtue of the introduction of the complementary event, taken in itself. Thus, ultimately what matters is not whether moral properties are objective in all the senses in which the term is used, but only that they are objective in all the senses in which it is relevant to so describe the property in light of a given aim.³²

From the dimensions of objectivity identified here we can see that various properties may be more and less objective *in toto* according to whether and how objective they are on these dimensions. At the objective extreme lie properties the existence of which do not depend on human responses at all (i.e. properties that are *not* response-dependent) such as being hard, cuboid, bouncy, and ‘in motion’.^{33/ 34} At the non-objective extreme lie properties that, for example, are less likely to survive hypothetical cases where humanity’s

³² The terms ‘objective’ and ‘non-objective’ are also used in other ways for properties: Richard Joyce discusses the possibility of talking about moral values as being objective in the sense that Michael Dummett defines ‘robust realism’ as the view that some properties are recognition-transcendent. Joyce, ‘Moral Anti-Realism’.

³³ Jackson and Pettit argue that all concepts are response-dependent, including the concept of hardness, and the lack of response-specificity is no bar to this: Jackson and Pettit, ‘Response-dependence without Tears’, 109–111. Notably, even if this is right, our focus is on properties.

³⁴ While motion only exists relative to a given location, it still exists without essentially depending on some response by us. Perhaps, for all this, a human response is required to identify the *location* that is the point of reference for motion. If so, then the property of being in motion may be slightly less objective than being hard, circular, and bouncy.

internal constitution changes, properties for which we are more willing to tolerate disagreement, and which have a narrow cosmological role. (Possible candidates include: being seductive, exasperating, chic, amusing, relaxing and spine-tingling).

How objective then are moral properties? We have already established that moral properties are essentially dependent on humans and their capacity to recognize them and this entails that such properties are not modally independent of humans and perhaps creatures relevantly similar to us (non-objective). That moral properties are so dependent on humans and their capacity to recognize them also means that their existence is dependent upon our being able to refer to subjects having a certain response to them in appropriate circumstances (non-objective). From Chapter 7 we may agree that moral properties entail reasons for action that are binding on rational creatures (objective). Although when fleshed out, Smith considers this sense of 'objectivity' to mean that the *same* reason is binding on all rational creatures were they to be in the same circumstances at the same time, and I have denied that this is necessarily so, at least on the basis of Smith's Internalism (non-objective).³⁵ We can also readily observe that moral properties do not correspond to a specific microphysical property (such as photons pulsating at a distinct frequency in the case of a colour property) or an identifiable range of such properties (such as the set of molecules responsible for eliciting the sensation of sourness) (non-objective).³⁶ Yet, moral properties clearly apply to physical acts and omissions which are, at least, modally independent of a third-party making a moral judgment (objective). Our unwillingness to consider that it may be that two agents

³⁵ See Chapter 4, III.

³⁶ According to Cornell Realists, specific moral properties, such as moral goodness, may correspond to distinct natural properties, such as happiness, or a group of properties. This is controversial, but even if true, happiness is not in turn elicited by a specific property or range of properties. It is potentially elicited by anything.

disagreeing about what morality requires are both right also favours objectivity, although such evidence is weak (objective).

On reflection it is difficult to identify anything that moral properties may explain aside from agents' moral judgments and responses to situations as they arise. Some academics argue that moral values are causally efficacious,³⁷ but they do not endorse the somewhat dubious contention that moral values causally affect properties or states of affair other than agents' moral judgments and their responses to them. Moral properties globally supervene on mental and physical natural properties, and not *vice versa* and I am yet to learn of an argument that they causally affect logical, mathematical, temporal and theistic facts. They causally affect historical and statistical facts, yet this is only ever via firstly affecting people's judgments and behaviour. Accordingly, without further evidence I agree with Wright that moral properties have a narrow cosmological role (non-objective).³⁸

C. Rigidity: Colours and Gustatory Properties

The last dimension of objectivity identified above is whether or not the moral value of agent *A* Φ -ing in *C* at *t* may change with a significant change to humanity's moral judgments to Φ -ing, in *C* at *t*. The existence of a moral property would not be required to explain that moral properties do not so change, yet if they do not, this provides a stronger case for the view that moral properties explain or partly explain the truth of correct moral judgments and responses.

On this point Wiggins explains that although, as he has argued, very often value properties can only be identified in virtue of the responses they elicit,

³⁷ For example: Oddie, *Value, Reality, and Desire*, 2, 3, 19–23, chap. 7.

³⁸ Wright, *Truth and Objectivity*, 197–8.

this does not mean that values may change with a significant change to humanity's judgments:

What if, by a sequence of minute shifts in our responses, an evil demon were to work us round to a point where we took what is actually evil to be good? Perhaps the demon might do this without our even noticing it. [...] But this is not an objection to sensible subjectivism. It would not follow from our not noticing the magnitude of the shift and everything that went with it that the very same thing that once told the presence of good was now fastened constitutively upon evil. For the subjectivism we have envisaged does not treat the response as a criterion, or even as an indicator. [...] it counts as nothing less than an act of judging a content; it is a judgment indispensably sustained by the perceptions and feelings and thoughts that are open to criticism that is based on norms that are open to criticism. It [our act of judging a content] is not that *by which we tell*. It is part of the telling itself.^{39/40}

Accordingly, Wiggins considers that the moral goodness or evilness of a given act can be held fixed according to its correct value status in the actual world now (or in the actual world at any time or most times in which human moral agents exist).

We can get a hold of this issue by considering how this dimension of objectivity applies to other properties. For example, on the face of it, colour properties are objective in this sense. Where we suppose that after overnight, worldwide surgery, what humans previously saw as the colour violet in external stimuli, their dreams, and imagination, came to be seen as orange, some writers intuitively consider that the extensions of violet and orange

³⁹ Wiggins, 'A Sensible Subjectivism?', 207–8 (emphasis in italics present). See also Mark Johnston's assertion that it is not an objection to claim that in many cases there are no standard perceivers or standard viewing conditions. Similarly, Firth argues that his account explains ethical statements so that it is not the case: 'that they would all be false by definition if there existed no experiencing subjects (past, present, or future)'. Johnston, 'Objectivity Refigured: Pragmatism Without Verificationism', 104–5; Firth, 'Ethical Absolutism and the Ideal Observer', 322.

⁴⁰ Blackburn and Gibbard also make this point within their own relativistic framework: Blackburn, *Ruling Passions*, 311; Gibbard, *Thinking How to Live*, 186.

remain as they were before the surgery.⁴¹ Peter Railton explains that this intuition can be captured by the assertion of (2) rather than (1) in the following two claims about red (instead of violet):

- (1) x is red = x is such as to elicit in normal humans (and in normal circumstances) the visual impression of redness.⁴²
- (2) x is red = x is such as to elicit in normal humans as they actually are [now] (and in actually normal circumstances [now]) the visual impression of redness.⁴³

‘Actually’ is included in (2) to qualify the relevant noun by referring to it as it exists in this world, compared to the first definition which is non-specific as to the relevant world.

In contrast to colour properties, however, Railton agrees with Sidney Shoemaker that, intuitively, we *deny* that gustatory properties exist so rigidly. If after overnight surgery, they argue, what previously tasted spicy to everyone, such as the Jamaican hot pepper, now tastes sour, the taste is now sour (and not still spicy).⁴⁴

So far as the details of these two hypothetical cases are spelt out, I agree with the intuitive results suggested. Yet this is not to say much at all. Indeed, such conclusions are misleading because the thought experiment so described generates different results according to how the following further details are spelt out:

- a) whether the subjects are aware of the change to their perceptual system after the surgery even though, on either account, they can still distinguish between different photons, molecules;

⁴¹ For example: Wright, *Truth and Objectivity*, 113; and perhaps: Railton, ‘Red, Bitter, Good’, 69.

⁴² Railton, ‘Red, Bitter, Good’, 68.

⁴³ Ibid., 69. See also: Wright, *Truth and Objectivity*, 114.

⁴⁴ Railton, ‘Red, Bitter, Good’, 76–7.

b) whether the ‘property of violet’ and the ‘property of spiciness’ refer to a disposition of the object or physical property to elicit a certain perception (e.g. looks violet, tastes sour) or whether the property in question refers to the perception itself;

ba) Moreover, if the former, whether the disposition is characterized as:

i) the higher-order property (or role property) of being such that the object or physical property elicits a certain perception, and being such that it elicits this response in virtue of certain lower-order properties (or realizer properties); or

ii) the lower-order properties of that object or physical property that may realize the higher-order property.

c) whether in conducting the thought experiment we do not acknowledge our own point of view as philosophers looking in on this world from the outside and an answer is instead given from the point of view of *those* humans who were subject to surgery or whether, in conducting the thought experiment, we *do* acknowledge our own point of view as philosophers looking in on this world from the outside, and an answer is given from our own point of view;

d) whether it is stipulated that the humans in question will or will not perceive the relevant sensation at some subsequent point in time or whether it is stipulated that this is unknown.

The upshot is that the thought experiment for each type of property can each be described in 36 relevantly different ways.⁴⁵ More importantly, if you consider each of the 36 cases for colour and gustatory properties, I think you shall realize that the intuitive results in each case are the same, or at least that, to the extent that a different intuitive result emerges, upon reflecting on

⁴⁵ Perhaps further relevant variables may also be identified, but this list suffices to make the point that it is misleading to provide a thought experiment about a change to human colour perception without attending to further details. Simon Blackburn makes a slightly similar point in regard to Mark Johnston’s contention that a shy and intuitive chameleon is green in the dark, even if he is always red in the light. Blackburn observes that we can regard the chameleon as either red or green, depending on how we exercise our imagination and how we understand the property of being green: Blackburn, ‘Circles, Finks, Smells and Biconditionals’, 265–7.

the relevant thought experiment for a colour and a gustatory property in turn, the rationalization of the intuitive result in one case, also makes sense in the corresponding case.

For example, suppose that after worldwide overnight surgery, unbeknownst to them, humans are no longer able to sense spiciness and all evidence that they could previously experience spiciness is erased, although they are still able to identify the various molecules that previously elicited this sensation. In this case, where, in the thought experiment, we *do* acknowledge the point of view of us philosophers conducting this thought experiment looking in on such a world (and hence we do consciously allow for an answer to be given from their point of view), and where it is stipulated that we *do not know* if humans or creatures with similar gustatory systems will be able to sense spiciness again, then where the property of spiciness is understood as ‘the property of perceiving spiciness’, I intuitively consider that the property does not exist, since, *ex hypothesi*, it presently cannot be so sensed. Furthermore, where these features of the thought experiment are retained, except that the property of spiciness is understood to be ‘the disposition in an object or molecules to elicit the sensation of spiciness’, in the form of the higher-order role property, then firstly I think that the object and molecules still exist: for example, I still think that the Jamaican hot pepper and the capsaicin and dihydrocapsaicin molecules in the pepper that were previously responsible for the sensation of spiciness still exist. Secondly, I think that beyond this, the dispositional property of spiciness itself still exists, at least for now. The results are different if it is instead stipulated that human subjects and like creatures will never experience spiciness again: in this case I deny that the property of spiciness as either the property of sensing spiciness or as a dispositional, higher-order role property in objects, for it will never elicit this response again.

In regard to colour, suppose that after worldwide overnight surgery, unbeknownst to them, humans are no longer able to perceive the colour violet and all evidence that they could previously experience violet is erased, although they are still able to identify and distinguish the photons pulsating in a wave-like manner every 400 nanometres and the combinations of other photons that previously elicited this perception. In this case, where, in the thought experiment, we *do* acknowledge the point of view of us philosophers conducting this thought experiment looking in on such a world (and hence we do allow for an answer to be given from their point of view), and where it is stipulated that we do not know if humans or creatures with similar visual systems will be able to sense the colour violet again, then where the property of violet is understood as 'the property of sensing violet', I intuitively consider that the property does not exist, since, *ex hypothesi*, it presently cannot be so sensed. Furthermore, where the features of the thought experiment remains the same, except that the property of violet is understood to be 'the disposition in an object or photons to elicit the sensation of violet', in the form of the higher-order role property, then firstly I think that the object and chemicals on its surface responsible for absorbing certain photons still exist: for example, I still think that the violet flower and the absorbent chemicals on the surface of its petals that were previously (indirectly) responsible for the sensation of violet still exists.⁴⁶ Secondly, I think that beyond this, the dispositional property of violet itself still exists, at least for now. The results are different if it is instead stipulated that human subjects and like creatures will never perceive the colour violet again: in this

⁴⁶ It is sometimes stated that an object's surface reflectance properties are responsible or partly responsible for eliciting a particular sensation of colour. Yet, so far as science explains colour experience, we perceive the colours of objects we perceive because of the *absence* of a type(s) of chemical(s) on the object's surface that would otherwise absorb the particular frequency of photons in question and because of the presence of types of chemicals that absorb photons of other frequencies. A preparedness to talk about surface reflectance properties is then appropriate so long as one endorses a Meinongian theory that counts nonexistent properties but I instead use the terms: surface absorbent properties, surface properties or chemicals responsible for absorbing photons and similar expressions.

case I deny that the property of violet exists as either the property of perceiving violet or as a dispositional, higher-order role property in objects, for objects and surface chemicals will never elicit this response again.

The difference that Shoemaker and Railton purport to identify is explained as a difference between understanding spiciness/ violet as **i)** the property of an object being perceived as spicy/ violet; compared to understanding it as **ii)** a disposition of the object or microphysical substrate to elicit this response (as either a role or realizer property) (variable **b**) of the thought experiment). When this difference is not expressly adverted to we have a tendency to conceive of colour properties as dispositional properties in the coloured object or photon itself whereas in the case of gustatory properties our tendency is to identify such properties with the subject's perception of the taste in question.

Shoemaker defends the distinction between colour and gustatory properties on the following basis, which Railton quotes with approval:

Our dominant interest in classifying things by flavor is our interest in having certain taste experiences and avoiding others, and not our interest in what such experiences tell us about other matters. With color it is the other way around; [...]⁴⁷

This explanation and justification of our different intuitive responses to the two surgical cases suggests that our main interest for gustatory properties is the 'taste experience', an interest that therefore heavily favours our understanding of gustatory properties as properties identified with the perceptive response. In contrast, the suggestion is that for colour perception our interest does not lie so much in the perception as it does in the information this provides us about other matters, an interests that therefore heavily favours our association of colour properties with some kind of

⁴⁷ Shoemaker, 'Self-Knowledge and "Inner Sense"', 302–3.

property in the object itself. However, once we are aware that colour and gustatory properties may be identified with either a certain response or a disposition in the object, we must ask for the justification for supposing that one view is correct over the other. The answer cannot derive from the thought experiment itself because in fact the thought experiment does not distinguish between the two possibilities of which we are now aware. If the thought experiment is amended such that the choice of property type is made clear in each case, and other relevant variables are held constant, the intuitive difference between the two cases disappears. Nor can the justification for favouring one view over the other justifiably derive from a pre-reflective interest in classifying properties a certain way, because insofar as the existence of that interest is based on pre-reflective linguistic evidence, this type of evidence is only of limited value.⁴⁸

Accordingly, while properties about an object's hardness, shape, weight and bounciness as well as the properties for colours and gustatory sensations may all be explained at least partly in virtue of microphysical properties, properties about hardness, shape, weight and bounciness can be distinguished from colour and gustatory properties insofar as the two latter properties do not necessarily survive cases in the same way, if at all, where humanity's internal constitution changes.

In turn, colour and gustatory properties can be distinguished from the response-dependent properties of being seductive, amusing, chic, relaxing and spine-tingling. Unlike properties of the former kind, for properties of the

⁴⁸ Where colour and gustatory properties diverge is in the degree to which we are willing to condone disagreement about our description of a certain sensation. To some extent we are willing to tolerate disagreement in regard to both descriptive colour and descriptive gustatory sensations, but we are more tolerant about the sensation and the degree to which something is spicy, sour, sweet, bitter among people, especially between adults and children. Further variability arises when we move to *evaluating* taste sensations—whether the taste is *umami* (delicious) or horrible, for example—yet such variability is also present among colour preferences.

latter kind, the evaluating response is far more important than the eliciting property. This is also partly reflected in the fact that the objects, states of affair and events that are amusing, chic, relaxing and spine-tingling cannot be classified with any kind of unity or consensus, other than their capacity to elicit the response in question—they are ‘shapeless’.⁴⁹ As a result, we are more readily inclined to explain our attribution of amusement to a certain situation in virtue of the agent *determining* that such a property exists, rather than in virtue of the agent recognizing that that property exists, and in particular, existing independently of her particular identification of it.

D. Rigidity: Moral Properties

Moral properties do not adequately fit into any of these three classes of properties in regard to this sense of objectivity. Rather, on this point I agree with Railton’s contention that in some cases moral properties survive hypothetical scenarios wherein humanity’s internal constitution is changed and in other cases they do not.^{50/ 51} Moral properties do not survive such cases where the moral value of performing a particular act is partly dependent on humanity’s internal constitution. For example, Railton supposes that in our actual world we intrinsically value a member of our kin over a stranger, such that, we would, for example, prefer to bestow a benefit

⁴⁹ Blackburn, *Ruling Passions*, 98. It may also be that there is no common denominator among molecules responsible for gustatory sensations. In any event, it is true that the class of physical features responsible for sweet or spicy sensations has definite boundaries and is far more restricted than physical features that may potentially be responsible for eliciting responses of: amusement, nausea, and embarrassment.

⁵⁰ Railton, ‘Red, Bitter, Good’, 83–4. See also: Lebar, ‘Three Dogmas of Response-Dependence’, secs. 6–9.

⁵¹ Interestingly, this may also be true of *colour* properties because, unlike the colours in the electromagnetic spectrum, pink, brown, and white cannot be observed in virtue of a set of photons of one frequency. Instead, they only ever arise from a combination of photons of different frequencies.

Indeed, it is also for this reason that it would be more difficult for Mary to *know* of (cf. imagine) other people’s experience of pink compared to red and other colours.

on a filial relation, rather than a stranger, where both are equally deserving. If all value terms were rigid designators then, as in the case of terms describing an object's shape, we would expect that favouring one's offspring in a certain way possesses intrinsic value even where humans' internal constitution changed, just as we believe that the beach ball remains spherical despite our internal constitution changing so as to visually perceive spherical objects as ellipsoids.

Yet for intrinsic value, Railton argues, this is not always so. Imagine, he writes, that with advances in medical technology human beings desire to reproduce by creating replicates of themselves.⁵² Among other things, a replicate does not quite look like her progenitor, she lacks her progenitor's memories and also any special intrinsic interest in her progenitor's life and the life of her biological relations even after having fully 'awakened' to life. Railton explains that replicates often go through the process of awakening with their kin, but that it is also not uncommon for such awakenings to occur with a 'host family' and the replicate herself does not care for one process over the other. In such a state of the world, Railton imagines that a replicate raised by a host family enters a taxi with a stranger on either side of him. Coincidentally, the stranger on his left is his progenitor, as he would learn if he struck up a conversation with him. Yet Railton argues that there is no more intrinsic value associated with the replicate striking up a conversation with the 'stranger' on his left than there is with the stranger on his right because replicates and their progenitors do not care about mere biological relatedness. Consequently, Railton argues, rigid designation of intrinsic value terms yields the wrong result in such a case.

Similarly, we may imagine a possible world that is exactly the same as our own with the sole exception that it is constitutive of human beings in this

⁵² This thought experiment comes from Railton: Railton, 'Red, Bitter, Good', 71–5.

world that they possess a propensity to believe (perhaps in virtue of a divine command or perhaps not) that ducks are sacred, especially the white-faced whistling duck, and accordingly humans in this world tend to believe that ducks should not be killed for food, perhaps even in cases of dire need. In this world we may imagine that a person's discretely selling the white-faced whistling duck for consumption is immoral and, given this we may doubt that the moral permissibility of such an act in our own world is rigidly designated.⁵³

Of course, from our discussion of the thought experiments on colour and gustatory sensations we know that certain further details in these two experiments must also be mentioned. Yet for some cases, where these details are introduced, the same intuitive results emerge and so for Railton's hypothetical and the hypothetical wherein ducks are sacred, moral properties behave in the same way as colour and gustatory properties.⁵⁴

In response to Railton's hypothetical, Brad Thompson argues that given the occurrence of replication in Railton's purported counterexample the facts about one's 'parents' and 'kin' in this imagined future differ from the facts about one's parents and kin in the earlier world such that it is not clear that the same intrinsic value of favouring one's kin applies in each case.⁵⁵ Hence, Thompson's concern is that Railton's thought experiment merely illustrates

⁵³ It may or may not be that killing ducks for consumption (generally and not in cases of dire need) is immoral in the actual world. Yet, whether or not this is true, it is not presently widely accepted and so the appropriate contrast arises.

⁵⁴ For example, the subjects may be aware of the change to their internal constitution from history books, we may be referring to moral rightness as a disposition elicited from a certain situation, the point of view of philosophers in the thought experiment may be acknowledged and it may be stipulated that replicates will never care about biological relatedness in the future. If we also suppose for clarity that two centuries have passed since there existed an ordinary human in such a world, then in these circumstances, I intuitively doubt that a dispositional moral property to care for one's biological relations exists.

⁵⁵ Thompson, 'Moral Value, Response-Dependence, and Rigid Designation', 81.

the point that moral values globally supervene on natural facts; it does not show that moral values do not always exist rigidly in the face of changes to humanity's internal constitution.

However, this concern does not take issue with the central point. All parties agree that a relevant difference between two factual scenarios may mean that different moral value properties arise for the same act. What is at issue is what may constitute a 'relevant' difference between factual scenarios where moral properties are thought to exist rigidly. In this regard, Railton's understanding of rigidity follows the analogy provided by colour terms. Hence, for Railton, maintaining that a property exists rigidly entails maintaining that it continues to exist, as it is, notwithstanding changes to human beings' internal constitution. This, in turn, means that Railton considers that if a property exists rigidly, a change to humanity's internal constitution will not constitute a relevant difference between two factual scenarios justifying one's assertion that the property in question differs in each case. Thompson must either: **a)** deny that intrinsic value properties always exist rigidly; or **b)** maintain that there is intrinsic value associated with the replicate conversing with his progenitor rather than the stranger even though neither party cares about biological relatedness; and that selling the white-faced whistling duck for consumption is not immoral in a world where it is constitutive of human beings that they tend to believe that ducks are sacred and ought not to be harmed.

There are however, some events for which we are less willing to be flexible in our application of moral rightness. Other things being equal, we do not consider that acts of whimsical cruelty, such as throwing stones at a swan, are morally right irrespective of what we came to feel and believe about this after worldwide, overnight surgery. In virtue of this, Railton distinguishes between cases where the change in human persona changes something that is '*partially constitutive* of a moral good' in question (such as valuing one's

kin and valuing ducks) and changes that are not.⁵⁶ Railton argues that the moral wrongness in acts of senseless cruelty lies in the torment to the victim in question (for example, swans) and hence not in something that could be changed by a change in humanity's persona. In light of the argument in the foregoing chapters, we can see that this contention is only partly right, but the qualification required here is presently unimportant. Rather, the point of note here is that in virtue of moral properties sometimes existing rigidly and sometimes not, Wiggins' conclusion on this point is in need of further justification or it remains in error. Moreover we can see that, in this sense of objectivity, moral properties do not fall in with either the properties about colour or gustation on the one hand, nor the properties of being seductive, amusing, chic, exasperating, relaxing or spine-tingling on the other.

E. Sensibility Theory Qualified

For all this, the non-rigid existence of some moral properties is merely one aspect of objectivity (albeit a seemingly important one) and accordingly, it does not strike at the heart of extension-reflecting theories. A proponent of Wiggins' theory could, for example, account for it by adopting one of two increasingly less robust positions. Firstly, she may agree that moral properties do not always exist rigidly and yet still maintain that, nevertheless, correct moral judgements are always extension-reflecting. Thus, on this first qualified version of Wiggins' theory, one will agree that in Railton's hypothetical future there is no intrinsic value associated with a replicate conversing with the stranger who is his progenitor (contrary to the present moral climate in the actual world) and it is agreed that, in the alternative hypothetical, it is immoral to sell the white-faced whistling duck for consumption (contrary to the present situation in the actual world). Yet in

⁵⁶ Railton, 'Red, Bitter, Good', 83–4. Blackburn also argues that morality should change according to our attitudes in some cases, but not others. Blackburn, *Ruling Passions*, 311–2.

each case one contends that the correct moral judgment is correct partly in virtue of the existence of a moral property that arises in each case. Furthermore, a proponent of this view may contend that there is nothing unusual about this result: it is just the implication of the theory that moral properties globally supervene on natural properties and facts, including facts about humanity's internal make-up.

This is a more plausible position than that espoused by McDowell and Wiggins. However, I set it aside here because, insofar as such a view considers that moral properties genuinely exist even where they do not do so rigidly, I wonder what consideration sufficiently justifies inferring the existence of the moral property when, at least on the face of things, less metaphysically committed views are equally viable. We have already seen that McDowell and Wiggins' arguments are unpersuasive on this front. In addition, we have seen that of the eight dimensions of objectivity identified, moral properties are only relatively objective in three: **1)** that physical conduct is modally independent of another's judgment of it; **2)** moral values entail reasons for action that are binding on rational creatures and **3)** that moral language has pretensions to objectivity. None of these dimensions is particularly relevant to proving that moral properties genuinely exist. The first two concern conduct and certain reasons binding all rational creatures respectively and not the moral property associated with it and we have discussed the limitations of linguistic evidence. Hence, where moral properties exist non-rigidly, the insistence that a moral judgment is right partly in virtue of the existence of a moral property does not enable us to explain anything more than what some extension-determining theories may be able to explain.

As a result, proponents of this view would do better by contending that moral judgments are extension-reflecting where moral properties exist rigidly and extension-determining where they exist non-rigidly (the second

less robust position available for proponents of extension-reflecting theories). However, as may be evident, such a theory faces the difficulty of justifying the existence of two different kinds of explanation for moral properties. A more elegant answer provides one kind of explanation for both rigidly and non-rigidly existing moral properties: a given moral property exists in light of a type of moral agent determining that property to exist. I therefore turn to consider the plausibility of a particular extension-determining theory.

II Judicious Agent Theory

More than one developed metaethical theory may be described as extension-determining. In the space remaining I want to explain why the source of moral values may plausibly be located in facts about what an ideal observer would judge or in facts about the outcome that would arise from the employment of a certain procedure.

Recall:

Morally Right: It is morally right for A to Φ in C at $t \leftrightarrow \Phi$ -ing is a morally relevant act & ($A+$ could recognize and be motivated to Φ in C at t for sufficient justifying reason R) & (δ)

where $A+$ denotes A reasoning as well as A could in C at t .

According to an ideal observer theory, covariation in truth in this biconditional is explained by virtue of the left side only ever obtaining *because* $A+$ could recognize and be motivated to Φ etc. in circumstances where Φ -ing is a morally relevant act. In short, moral rightness follows from what $A+$ could recognize and be motivated to do: the right side has explanatory priority.

Ultimately, there is no relevant difference between this kind of theory and metaethical constructivism and so, we could equally explain why moral rightness can plausibly be explained by the employment of a procedure involving one firstly putting oneself in the position of an ideal agent. But I need not defend this point here.⁵⁷ Instead I turn to various objections put to ideal observer theories, some of which are also put to constructivist theories.

A. Objections

A trite objection to ideal observer theories and metaethical constructivism consists in one fork of the Euthyphro dilemma: there is insufficient justification to believe that what an ideal observer judges to be morally right or wrong is actually right or wrong.⁵⁸ Thus, for example, Sharon Street argues that, analogously, no matter how detailed our account of an ideal baseball umpire is and the circumstances in which she renders her judgment, there may always be a significant divergence between when we intuitively think a certain player is safe and what the umpire judges.⁵⁹ In a related vein,

⁵⁷ For statements and defences of metaethical constructivism see especially the work of Sharon Street, as well as James Lenman and Christine Korsgaard's 'procedural realism': Street, 'What Is Constructivism in Ethics and Metaethics?', 366; Street, 'Constructivism about Reasons', 223; Lenman, 'Humean Constructivism in Moral Theory'; Korsgaard, *The Sources of Normativity*, 36; Korsgaard, 'Realism and Constructivism in Twentieth-Century Moral Philosophy'. David Enoch makes the point that ideal observer theory and metaethical constructivism are not relevantly different, both on standard definitions of the theory and Street's slightly different definition: Enoch, 'Can There Be a Global, Interesting, Coherent Constructivism about Practical Reason?', 328–9. On reflection, it also appears that this is so. The ideal observer or agent employs a certain procedure of reasoning. One need only add to the start of this procedure the steps required to become an 'ideal agent', according to a given ideal observer theory, to be able to say that moral values are grounded in a certain procedure, instead of what the ideal observer would or could conclude.

⁵⁸ Wiggins, 'A Sensible Subjectivism?', 192; Wright, 'Moral Values, Projection and Secondary Qualities', 23–4; Street, 'What Is Constructivism in Ethics and Metaethics?', 37; Gibbard, *Wise Choices, Apt Feelings*, 18–22; Skorupski, *The Domain of Reasons*, 447; Joyce, 'The Accidental Error Theorist', 166. Russ Shafer-Landau puts the Euthyphro objection to metaethical constructivist theories: Shafer-Landau, *Moral Realism*, 41; Shafer-Landau, 'Replies to Critics', 315–6.

⁵⁹ Street, 'What Is Constructivism in Ethics and Metaethics?', 37.

Wiggins argues that we have an insufficient grasp and an insufficient account of our actual grasp of what constitutes a good critic or judge.⁶⁰

It might be thought that Street's criticism applies even more readily to the ideal observer theory in Morally Right given that *A+* is only idealized in the sense that she reasons correctly. That is, she does not necessarily have no false beliefs and all relevant true beliefs and as a result there are more ways she can be in error. Yet, in effect this worry has already been answered in our preceding discussion of internal reasons, Chapters 4-6, where one of the main bases for the argument that there can be no pure external reasons (in the sense espoused by Immodest Externalism) is that there is no non-arbitrary distinction between cases where: **a)** it is stipulated that it is physically impossible for a given agent, *A*, to Φ in *C* at *t* in a deterministic sense; and cases where: **b)** it is stipulated that it is logically impossible for a given agent, *A*, to Φ in *C* at *t*. In both cases *A* will not be able to perform the relevant act and I accordingly argued that we should give up on pure external reasons. Since moral values entail sufficient justifying reasons, the same point applies to moral cases as a subset of all cases considered earlier. The analogy to the umpire collapses insofar as the umpire is asked to make a judgment about theoretical reasons we have, compared to the ideal observer who must make judgments about practical reasons moral agents have.⁶¹

Accordingly, to the extent that it is stipulated that *A+* could not recognize or be motivated to act on a certain consideration, she does not have a justifying reason to perform that act, and correspondingly, there is no moral value associated with her doing so. The thought that counterexamples readily apply to the ideal observer under consideration therefore relies on an intuition arising in each such example that could only be put forward by disregarding our discussion in the last six chapters. Indeed, in virtue of this,

⁶⁰ Wiggins, 'A Sensible Subjectivism?', 192.

⁶¹ See Chapter 6, IV for a brief discussion of this distinction.

such intuitions have very little probative value at all in this context, for the entire point of developing a metaethical theory of any kind is to provide a reflective account of moral values in lieu of such intuitions. It would defeat the point of pursuing such a task if after providing such an account based on detailed consideration of various intuitions, among other things, one were to simply stop, throwing one's hands up in futility upon recognizing that one's theory is inconsistent with a pre-reflective intuition that the justification for the theory has already explained.

Wiggins' criticism has slightly more merit. My account of *A*⁺ is relatively minimal and avoids the difficulties associated with imagining what someone with no false beliefs and all true beliefs could judge that she should do. The person who is *A* is not a counterfactual artifice to any degree: she is just *A*, reasoning as well as *A* possibly could in the circumstances and at the time in question. It may still be put to such a version of ideal observer theory that we have an insufficient grasp of what 'reasoning as well as *A* possibly could' consists in. Yet, I doubt that the result of identifying the truth conditions of moral requirements in such an ideal agent will be indeterminate in all cases and, even if they were, if the aim is to identify such truth conditions, then the fact that specific results are indeterminate on such a theory does not mean that the theory is false so the consideration turns out to be beside the point.⁶²

A second and close objection to ideal observer theories that Street puts forward is what Mark Schroeder describes in another context as the 'Wrong Place objection'.⁶³ According to Street, just as it does not seem right to consider that a player's being safe is constituted by facts about the ideal

⁶² Indeed, as Williams argues, it is more realistic because of this: Williams, 'Internal and External Reasons', 110[3]; Williams, 'Internal Reasons and the Obscurity of Blame', 38[2].

⁶³ Schroeder, *Slaves of the Passions*, 37–8. For example, Thomas Hill Jr. argues that, contrary to one version of Kant's categorical imperative, moral wrongness does not exist in an inconsistency. Hill Jr., 'Kantian Normative Ethics', 488.

umpire judging this to be the case, it also does not seem right to consider that an agent performing the morally right act is constituted by facts about the ideal observer judging this to be the case.⁶⁴ Somewhat similarly, in regard to Hume's projectivism, Wiggins argues that it is 'ludicrous' to suggest that a responsible judge of what is funny would firstly determine if a true judge would find *x* funny and then, supposing she does so, eschew stating 'gild or stain' *x*, or 'project' upon *x*.⁶⁵ Rather, Wiggins supposes that there is something in the object that is made for the sentiment it would occasion in a qualified judge. The criticism may be adapted and put to an ideal observer theory with the suggestion that it is false that values derive from what the ideal observer judges to be the case.

We have already seen that Street's analogy is false. Furthermore the weakness of both objections is that they rely on pre-reflective intuitions. After appreciating that an ideal observer theory can account for the relevant senses in which moral values are objective, there is nothing odd in supposing that moral values are so located. Indeed, this was one lesson derived from Moore's OQA (Complex Property). Wiggins' objection may then be put to metaethicists themselves by stating that if one were a proponent of ideal observer theory, then in one's non-philosophical life a responsible judge of moral values would have to try to identify what the ideal agent would consider to be morally right in the relevant circumstances and then subsequently be willing to expressly explain that a certain act is morally right for her to perform because she judges that this is the right thing to do when she reasons as well as she is possibly able. Yet, as a particular ideal observer theory, compared to projectivist theory, this theory does not appear ludicrous but rather it both appears to hit the mark perfectly and does hit the mark perfectly. For what more can we ask of her in *C* at *t*?

⁶⁴ Street, 'What Is Constructivism in Ethics and Metaethics?', 372–3.

⁶⁵ Wiggins, 'A Sensible Subjectivism?', 193–4.

Simon Blackburn objects to biconditional analyses of moral evaluations for different reasons.⁶⁶ Blackburn argues that where the aim of the analysis is to understand what it is *to think* in ethical terms or what it is to give an ethical judgment, the analysis faces a problem of regression and a dilemma. He acknowledges that analyses avoid these problems if the aim in giving an analysis is not to provide an understanding of what it is to give an ethical judgment but instead only to identify the truth conditions for such judgments. However, he maintains that we want an account of both: an account of what we are doing in giving ethical judgments and an account of their truth conditions.

So far I agree. As the dilemma he explicates shows, ethical thought is *sui generis* and incapable of being analysed as thoughts about natural or empirical phenomena. Furthermore, I agree we want an account of both the targets he identifies—why not? Although, to this Blackburn adds that if the analysis is only intended to give truth conditions, then there is a certain lack of harmony between the right-hand side expression and the agent's (expressive, compared to, reportive) response. Thus, he explains, I might accept that what it is to judge that something is boring is 'tending to make people fall asleep on contemplating it' but then there is a certain lack of harmony between this response and the analysis that purports to identify the truth condition for judging that something is boring:

Boring: X is boring \equiv I judge that X is boring.

The lack of harmony arises because I can find something boring and fall asleep in response to it without judging that I find it boring. At its heart then, the concern for a lack of harmony is a concern for there being a type of counterexample to the analysis of the truth conditions of 'X is boring' and, so

⁶⁶ Blackburn, 'Circles, Finks, Smells and Biconditionals', 270–4; Blackburn, *Ruling Passions*, 104–119.

the objection will maintain, a type of counterexample to an analysis of the truth conditions of moral evaluations that is stated in the same form.

This argument is less persuasive and, accordingly, I am not convinced that an account of what it is to make an ethical judgment and the truth conditions of moral evaluations must be so closely related. We can see this when we try to adapt the analogy so as to assess Morally Right. Firstly, the right-hand side of Morally Right does not state 'A+ judges that Φ -ing in C at t is right': it states, in part, that 'A+ *could* recognize and be motivated to Φ in C at t for sufficient justifying reason R'. This analysis also refers to the agent in the third person and not the first person but, as Blackburn observes, the lack of harmony between one tending to fall asleep and one judging that 'X is boring' arises in both the first and third person, so I set this difference aside. The appropriate analogy must, however, account for the difference in modality:

Boring*: X is boring \equiv A+ *could* judge that X is boring

This amendment makes a difference. Boring* is not necessarily inconsistent with X inducing A+ to fall asleep without A+ judging 'X is boring'. For all that we know A+ could have formed the judgment without falling asleep, before falling asleep or even *after* waking up. Of course, the example may be modified to account for this. Suppose, for example, that unbeknownst to her, Lawan finds staring at small Paris green objects so boring that, due to a quirk in her brain chemistry, it is not even possible for her to judge that this is boring before falling asleep—she falls asleep rather quickly. Suppose also that she is unable to make this judgment at a later time. Perhaps, in a story resembling Sleeping Beauty, for example, a witch tricks Lawan into staring

at a small Paris green pendant and successfully ensures that she does not wake, ever after.⁶⁷

Even if we accept such a counterexample to Boring*, we can see that counterexamples of the same kind do not threaten analyses of moral evaluations in the form of Morally Right. We have already acknowledged that moral facts may give moral agents justifying reasons to act where the content of the reason itself does not refer to the act being one's moral duty. So to maintain an analogous counterexample for moral evaluations we must suppose that, for a certain act, Φ -ing, that is morally right, one Φ s *without* thinking or judging that she has a justifying reason to Φ at all or without thinking or judging that she has a justifying reason of the right kind (where the right kind does not necessarily mean that the content of the reason refers to a moral duty). Yet, morality only regulates actual and possible acts and omissions that are: **a)** intentional (including intentional acts where one is also reckless as to the possibility of other consequences); and **b)** physically possible. And, *judging* that 'one should Φ in C for justifying reason R' at some point before actually Φ -ing is a condition of it being possible to intentionally Φ . So, if the counterexample in question stipulates that it is not rationally possible for A to judge that she has justifying reason R to Φ in C (where R is a justifying reason of the right kind) before Φ -ing, then Φ -ing cannot be morally required from the outset, and the counterexample falls by the wayside.

This is also true where (or if) morality requires an agent to react instinctively. Perhaps a soldier trained to react instinctively (break a hold as soon as she is grabbed; not panic; not hesitate to kill in certain conditions etc.) may be

⁶⁷ One may argue that the appropriate sense of 'could' here is one in which A+ could judge that X is boring if, in a situation resembling Sleepy Beauty such as that just described, she would/ could, *per impossible*, judge that X is boring upon being shown a vision of the events she went through. Yet I need not go so far in considering the illustration.

morally required to react in this way in actual combat given that not doing so may endanger the lives of other soldiers and civilians. Furthermore, it may be objected, such instinctive reactions do not involve the agent making a judgement (consciously or on some subconscious level) that she should so act—she just does so. Yet, the judgment in such cases is accordingly already made in response to imagined circumstances of the relevant kind in the soldier's training. And if the reaction is morally required, some such training must have occurred for it to be plausible to associate responsibility with the response.

In the case of Morally Right then, no lack of harmony arises between the analysis of the truth conditions for moral evaluations and agents' moral responses, and hence, there is no problem that the analysis only aims to identify the truth conditions for such evaluations.

The upshot is that the truth of Morally Right is plausibly explained by at least one extension-determining theory, what I call: Judicious Agent Theory. Moral properties derive from what a given moral agent could judge that she has a sufficient justifying reason to do in morally relevant circumstances where she reasons as well as she could in those circumstances at that time. It is in light of moral facts being explained by the agent's actual capacities, that the theory's title emphasizes the *agent* and not an *observer*. It also omits reference to the agent being *ideal* given that this often implies that the agent (or observer) has full information and reasons correctly, even if she is not, for other reasons, perfectly ideal. Describing the agent as *rational* is preferable, but may imply that moral facts are reducible to facts about rationality in itself, that is, reducible to facts about rationality extricated from a moral agent's particular exercise of it (Moral Rationalism). In contrast, a *judicious* agent is one who 'has or exercises sound judgment' where by 'sound' I mean

‘resulting from correct reasoning’, in the sense of correctness we have been discussing.⁶⁸

Despite this, as concluded in Chapter 3, moral reasons are non-contingently agent-neutral insofar as the appropriate sub-group of rational agents for moral reasons only ever comprises moral agents who could recognize and act for the particular reason at the relevant time. In virtue of this, Judicious Agent Theory is a version of Moral Realism: an agent-neutral relation of dependence obtains between moral facts and moral agents.

Yet it is not a Naturalistic Moral Realism: it is not predominantly concerned with actually locating moral and morally relevant facts in the world by defining them in a way that allows for their identification via the scientific method. According to Judicious Agent Theory, moral facts are to be located by determining what a given moral agent *A* has sufficient reason to do in morally relevant circumstances when reasoning as well as possible in those circumstances at that time. The process of testing predications from hypotheses by replicable experiments (perhaps surveys in particular) may allow us to identify what *A* believes about her current circumstances and the world at large, as well as what is important to her and what she wants. But, in itself, this will not always warrant a conclusion about ‘what she could believe she has sufficient reason to do, in certain circumstances, if she reasoned as well as she could in those circumstances at a certain time’: forming this conclusion will often, if not always, require independent reflection rather than replicable experiments. It may also require knowledge of the possibility and impossibility of certain kinds of non-rational intervention which, in turn, may require knowing what some other rational agent could (rationally or non-rationally) have done in the circumstances in

⁶⁸ For an explanation of this sense see: Chapter 4, III.

question. Unless one stipulates to the possibility or impossibility of these events, the scientific method falls short on this front for the same reason.

Yet nothing rests on whether Judicious Agent Theory is a version of Moral Realism or not, Non-Naturalism or not. One's unwillingness to countenance Judicious Agent Theory as a version of moral realism would not justify the slightest facial gesture. Indeed, writers already maintaining that moral realism requires, more robustly, that a moral fact obtain that is independent of moral agents will deny that Judicious Agent Theory is a realist theory. The same point applies to its Non-Naturalist standing. It is the substance of the theory that matters, particularly in this regard, the profession that moral facts need not extend beyond rational possibility for an individual relative to a certain point in time. And, indeed, this result is not that surprising: it is just a reflection of the thought that morality is thoroughly grounded in truths about the cognitive and physical capacities and limitations of human, moral agency and the inherent complexity of circumstances that arise in life.

Bibliography

- ADAMS, ROBERT MERRIHEW. *Finite and Infinite Goods: A Framework for Ethics*. New York: Oxford University Press, 1999. Oxford Scholarship Online.
- ALM, DAVID. 'An Argument for Agent-Neutral Value'. *Ratio: An International Journal of Analytic Philosophy* 20, no. 3 (September 2007): 249–263.
- ALSTON, WILLIAM P. 'Introduction'. In *Realism and Antirealism*, edited by William P. Alston, 1–9. Ithaca: Cornell University Press, 2002.
- ALTMAN, ANDREW. 'Breathing Life into a Dead Argument: G.E. Moore and the Open Question Argument'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 117, no. 3 (2004): 395–408.
- ANSCOMBE, G. E. M. *Intention*. 2nd ed. Cambridge, MA: Harvard University Press, 1957/ 2000.
- ARISTOTLE. *The Nicomachean Ethics*. London: Penguin Classics, 322 B.C./ 2004.
- AUDI, ROBERT. 'Moral Judgement and Reasons for Action'. In *Ethics and Practical Reason*, edited by Garrett Cullity and Berys Gaut, 125–159. Oxford: Clarendon Press, 1997.
- BAIOCCHI, TALIA. 'N.Y. Takes a New Look at California Wines'. *San Francisco Chronicle*, November 7, 2010. Accessed August 30, 2011.
<http://www.sfgate.com/cgi-bin/article.cgi?f=/c/a/2010/11/05/FD0S1G6MQS.DTL>.
- BALDWIN, THOMAS. 'A Hundred Years of *Principia Ethica*'. *Ethica: An International Journal for Moral Philosophy* 2, no. 1 (June 2003): 3–13.
- . 'Editor's Introduction'. In *Principia Ethica*, edited by Thomas Baldwin, ix–xxxvii. Cambridge: Cambridge University Press, 1993.
- . *G.E. Moore*. London: Routledge, 1990.
- . 'Logic and Philosophical Analysis'. In *The Cambridge History of Philosophy 1870–1945*, 417–423, 2003.
- . 'The Indefinability of Good'. *Journal of Value Inquiry* 37, no. 3 (2003): 313–328.
- BALL, STEPHEN W. 'Linguistic Intuitions and Varieties of Ethical Naturalism'. *Philosophy and Phenomenological Research* 51, no. 1 (March 1991): 1–38.
- . 'Reductionism in Ethics and Science: A Contemporary Look at G.E. Moore's Open-Question Argument'. *American Philosophical Quarterly* 25, no. 3 (1988): 197–213.
- BARRY, MELISSA. 'Realism, Rational Action, and the Humean Theory of Motivation'. *Ethical Theory and Moral Practice: An International Forum* 10, no. 3 (2007): 231–242.
- BEAULIEU, GERALD. 'Meta-Ethical Rationalism and the Amoralist Challenge: An Externalist Response to Michael Smith's Reliability Argument'. *Dialogue: Canadian Philosophical Review* 46, no. 4 (Fall 2007): 751–760.
- BEDKE, MATTHEW S. 'Practical Reasons, Practical Rationality, Practical Wisdom'. *Ethical Theory and Moral Practice: An International Forum* 11, no. 1 (February 2008): 85–111.
- BENN, S. I., and G. F. GAUS. 'Practical Rationality and Commitment'. *American Philosophical Quarterly* 23, no. 3 (July 1986): 255–266.
- BLACKBURN, SIMON. 'Circles, Finks, Smells and Biconditionals'. *Nous-Supplement: Philosophical Perspectives* 7, Language and Logic, no. 1 (1993): 259–279.
- . 'Errors and the Phenomenology of Value'. In *Morality and Objectivity*, edited by Ted Honderich, 1–22. London: Routledge & Kegan Paul, 1985.

- . 'Practical Tortoise Raising'. *Mind: A Quarterly Review of Philosophy* 104, no. 416 (October 1995): 695–711.
- . *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Clarendon Press, 1998.
- . 'The Majesty of Reason'. *Philosophy* 85 (2010): 5–27.
- BLOOMFIELD, PAUL. *Moral Reality*. New York: Oxford University Press, 2004.
- . 'Opening Questions, Following Rules'. In *Metaethics after Moore*, edited by Terence Horgan, 169–189. Oxford: Clarendon Press, 2006.
- BOGHOSSIAN, PAUL, and CHRISTOPHER PEACOCKE. 'Introduction'. In *New Essays on the A Priori*, 1–10. New York: Oxford University Press, 2001.
- BOND, E. J. *Reason and Value*. Cambridge: Cambridge University Press, 1983.
- BOYD, RICHARD. 'How to Be a Moral Realist'. In *Essays on Moral Realism*, 181–228. Ithaca: Cornell University Press, 1988.
- BRATMAN, MICHAEL E. 'Cognitivism about Practical Reason'. *Ethics: An International Journal of Social* 102, no. 1 (October 1991): 117–128.
- BREAKOUTBAND. 'We All Have a Reason to Celebrate! Lady Gaga's...'. May 2011. Accessed August 30, 2011.
<http://breakoutband.tumblr.com/post/5610125711/we-all-have-a-reason-to-celebrate-lady-gagas>.
- BRINK, DAVID O. 'A Puzzle About the Rational Authority of Morality'. *Philosophical Perspectives* 6 (1992): 1–26.
- . 'Moral Motivation'. *Ethics: An International Journal of Social* 108, no. 1 (1997): 4–32.
- . *Moral Realism and the Foundations of Ethics*. Cambridge University Press, 1989.
- . 'Moral Realism and the Sceptical Arguments from Disagreement and Queerness'. *Australasian Journal of Philosophy* 62 (1984): 111–125.
- . 'Realism, Naturalism and Moral Semantics'. *Social Philosophy and Policy* 18 (2001): 154–76.
- . 'Utilitarian Morality and the Personal Point of View'. *Journal of Philosophy* 83 (August 1986): 417–438.
- BROCK, STUART, and EDWIN MARES. *Realism and Anti-Realism*. Stocksfield: Acumen Publishing, 2007.
- BROOME, JOHN. 'Reasons and Motivation II'. *Proceedings of the Aristotelian Society, Supplementary Volume* 71, no. 1 (1997): 131–146.
- BROWER, BRUCE W. 'Dispositional Ethical Realism'. *Ethics: An International Journal of Social, Political, and Legal Philosophy* 103, no. 2 (1993): 221–249.
- BRUENING, WILLIAM H. *The Is-Ought Problem: Its History, Analysis, and Dissolution*. Washington: University Press of America, 1978.
- BUTCHVAROV, PANAYOT. 'Saying and Showing the Good'. In *Time and Ethics: Essays at the Intersection*, edited by Heather Dyke, 137–158. Dordrecht: Kluwer Academic Publishers, 2003.
- CARROLL, LEWIS. 'What the Tortoise Said to Achilles'. *Mind: A Quarterly Review of Philosophy* 104, no. 416 (1895/ October 1995): 691–693.
- CAWS, PETER. 'Ethics and Temporality: *When Are Moral Propositions True?' In *Time and Ethics: Essays at the Intersection*, edited by Heather Dyke, 99–114. Dordrecht: Kluwer Academic Publishers, 2003.
- CHALMERS, DAVID J. 'Facing Up to the Problem of Consciousness'. *Journal of Consciousness Studies* 2, no. 3 (1995): 200–19.
- . 'Phenomenal Concepts and the Knowledge Argument'. In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge*

- Argument*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 269–298. Cambridge, MA: Bradford Book, MIT Press, 2004.
- CLARK, ERNEST, and MARIA CLARK. 'What Is Goodness?' *Australasian Journal of Psychology and Philosophy* 19, no. 1 (1941): 144–64.
- COHON, RACHEL. 'Are External Reasons Impossible?' *Ethics: An International Journal of Social, Political, and Legal Philosophy* 96 (1986): 545–556.
- COLEMAN, MARY CLAYTON. 'Directions of Fit and the Humean Theory of Motivation'. *Australasian Journal of Philosophy* 86, no. 1 (March 2008): 127–139.
- COOKE, NATHALIE. *Margaret Atwood: A Critical Companion*. United States: Greenwood Publishing Group, 2004.
- COPP, DAVID. 'A Skeptical Challenge to Moral Non-Naturalism and a Defense of Constructivist Naturalism'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 126, no. 2 (2005): 269–283.
- . 'Moral Naturalism and Three Grades of Normativity'. In *Morality in a Natural World: Selected Essays in Metaethics*. Cambridge; New York: Cambridge University Press, 2004/ 2007.
- . 'The Normativity of Self-Grounded Reason'. In *Morality in a Natural World: Selected Essays in Metaethics*, edited by David Copp, 309–353. Cambridge: Cambridge University Press, 2005/ 2007.
- . 'Why Naturalism?' In *Morality in a Natural World: Selected Essays in Metaethics*, 33–54. Cambridge; New York: Cambridge University Press, 2003/ 2007.
- COPP, DAVID, and D. SOBEL. 'Desires, Motives, and Reasons: Scanlon's Rationalistic Moral Psychology'. *Social Theory and Practice* 28 (2002): 243–276.
- CRISP, ROGER. 'Naturalism and Non-Naturalism in Ethics'. In *Identity, Truth and Value*, edited by Savina Lovibond and S.G. Williams, 113–129. Aristotle Society Series 16. Malden, MA: Blackwell Publishers, 1996.
- . *Reasons and the Good*. Oxford: Clarendon Press, 2006.
- CULLITY, GARRETT. *The Moral Demands of Affluence*. Oxford: Oxford University Press, 2004.
- CUNEO, TERENCE. 'Are Moral Qualities Response-Dependent?' *Noûs* 35, no. 4 (December 2001): 569–591.
- . 'Recent Faces of Moral Nonnaturalism'. *Philosophy Compass* 2, no. 6 (2007): 850–879.
- . *The Normative Web: An Argument for Moral Realism*. Oxford: Oxford University Press, 2007. Oxford Scholarship Online.
- DALY, CHRIS. *An Introduction to Philosophical Methods*. Canada: Broadview Press, 2010.
- DANCY, JONATHAN. 'Agent-Relativity -- The Very Idea'. In *Value, Welfare, and Morality*, edited by R. G. Frey and Christopher W. Morris, 233–51. New York: Cambridge University Press, 1993.
- . *Ethics Without Principles*. Oxford: Clarendon Press, 2004.
- . 'Moral Particularism'. Edited by Edward N. Zalta. *Stanford Encyclopedia of Philosophy*, Spring, 2009.
<http://plato.stanford.edu/archives/spr2009/entries/moral-particularism/>.
- . *Practical Reality*. Oxford; New York: Oxford University Press, 2000.
- . 'Two Conceptions of Moral Realism I'. *Aristotelian Society: Supplementary Volume* 60, no. 1 (1986): 167–188.
- . 'What Do Reasons Do?' In *Metaethics after Moore*, edited by Terry Horgan and Mark Timmons, 39–59. Oxford: Clarendon Press, 2006.
- DARWALL, STEPHEN L. *Impartial Reason*. Ithaca: Cornell University Press, 1983.

- . 'Internalism and Agency'. *Nous-Supplement: Philosophical Perspectives* 6, Ethics, no. 1 (1992): 155–74.
- . 'Moore, Normativity, and Intrinsic Value'. *Ethics: An International Journal of Social* 113, no. 3 (2003): 468–489.
- DARWALL, STEPHEN L., ALLAN GIBBARD, and PETER RAILTON. 'Toward *Fin de siècle* Ethics: Some Trends'. *Philosophical Review* 101, no. 1 (1992): 115–89.
- DAVIDSON, DONALD. 'Appendix to Essay 10: Belief and the Basis of Meaning (1974)'. In *Inquiries into Truth and Interpretation*, 281–6. 2nd ed. New York: Oxford University Press, 2001.
- . 'Belief and the Basis of Meaning'. In *Inquiries into Truth and Interpretation*, 141–54. 2nd ed. New York: Oxford University Press, 1974.
- DE BRIGARD, FELIPE. 'If You Like It, Does It Matter If It's Real?' *Philosophical Psychology* 23, no. 1 (February 2010): 43–57.
- DENNETT, DANIEL C. "'Epiphenomenal" Qualia?' In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 59–68. Cambridge, MA: Bradford Book, MIT Press, 1991/2004.
- . 'Real Patterns'. *Journal of Philosophy* 88, no. 1 (1991): 27–51.
- DORSEY, DALE. 'Weak Anti-Rationalism and the Demands of Morality'. *Noûs* 46, no. 1 (2012): 1–23.
- DUMAS, ALEXANDRE. *The Count of Monte Cristo*. Great Britain: Wordsworth Editions, 1844-6/1997.
- DURRANT, R. G. 'The Identity of Properties and the Definition of "Good"'. *Australasian Journal of Philosophy* 48, no. 3 (1970): 360–1.
- DYKE, HEATHER. 'What Moral Realism can Learn from the Philosophy of Time'. In *Time and Ethics: Essays at the Intersection*, edited by Heather Dyke, 11–25. Dordrecht: Kluwer Academic Publishers, 2003.
- ENOCH, DAVID. 'Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action'. *Philosophical Review* 115, no. 2 (April 2006): 169–198.
- . 'An Outline for an Argument for Robust Metanormative Realism'. In *Oxford Studies in Metaethics vol. 2*, edited by Russ Shafer-Landau, 21–50. Oxford: Oxford University Press, 2007.
- . 'Can There Be a Global, Interesting, Coherent Constructivism about Practical Reason?' *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action* 12, no. 3 (September 2009): 319–339.
- FELDMAN, FRED. 'The Open Question Argument: What It Isn't; and What It Is'. *Nous-Supplement: Philosophical Issues* 15, Normativity, no. 1 (2005): 22–43.
- FINE, KIT. 'Essence and Modality: The Second Philosophical Perspectives Lecture'. *Nous-Supplement: Philosophical Perspectives* 8, Logic and Language, no. 1 (1994): 1–16.
- FINLAY, STEPHEN. 'Errors upon Errors: A Reply to Joyce'. *Australasian Journal of Philosophy* 89, no. 3 (2011): 535–547.
- . 'Four Faces of Moral Realism'. *Philosophy Compass* 2, no. 6 (2007): 820–849.
- . 'Responding to Normativity'. In *Oxford Studies in Metaethics vol. 2*, edited by Russ Shafer-Landau. Oxford: Oxford University Press, 2007.
- . 'The Error in the Error Theory'. *Australasian Journal of Philosophy* 86, no. 3 (September 2008): 347–369.
- . 'The Obscurity of Internal Reasons'. *Philosophers' Imprint* 9, no. 7 (July 2009): 1–22.
- . 'The Reasons That Matter'. *Australasian Journal of Philosophy* 84, no. 1 (March 2006): 1–20.

- FINLAY, STEPHEN, and MARK SCHROEDER. 'Reasons for Action: Internal vs. External'. Edited by Edward N. Zalta. *Stanford Encyclopedia of Philosophy*, Fall, 2008. <http://plato.stanford.edu/archives/fall2008/entries/reasons-internal-external/>.
- FIRTH, RODERICK. 'Ethical Absolutism and the Ideal Observer'. *Philosophy and Phenomenological Research* 12 (March 1952): 317–345.
- FITZPATRICK, WILLIAM J. 'Recent Work on Ethical Realism'. *Analysis* 69, no. 4 (2009): 746–760.
- . 'Robust Ethical Realism, Non-naturalism, and Normativity'. In *Oxford Studies in Metaethics vol. 3*, edited by Russ Shafer-Landau, 159–205. Oxford: Oxford University Press, 2008.
- FODOR, JERRY. *Psychosemantics*. Cambridge, MA: MIT Press, 1987.
- FOOT, PHILIPPA. *Natural Goodness*. Oxford: Oxford University Press, 2003.
- FRAMARIN, CHRISTOPHER G. 'Motivation-Encompassing Attitudes'. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action* 11, no. 2 (2008): 121–130.
- FRANK, ROBERT H. *The Economic Naturalist: Why Economics Explains Almost Everything*. Reprint. Great Britain: Virgin Books, 2007.
- FRANKENA, WILLIAM K. 'The Naturalistic Fallacy'. *Mind: A Quarterly Review of Philosophy* 48, no. 192 (1939): 464–77.
- FRANKFURT, HARRY. 'Rationalism in Ethics'. In *Autonomes Handeln: Beiträge zur Philosophie von Harry G. Frankfurt*, edited by Monika Betzler, 259–273. *Autonomes Handeln: Beiträge zur Philosophie von Harry G. Frankfurt*, 2000.
- . *Taking Ourselves Seriously and Getting It Right*. Stanford: Stanford University Press, 2006.
- FREGE, GOTTLÖB. 'On Sense and Reference'. In *Translations from the Philosophical Writings of Gottlob Frege*, edited by Peter Geach and Max Black, translated by Max Black, 56–78. 2nd ed. Oxford: Basil Blackwell, 1892/ 1977.
- FUMERTON, RICHARD. 'Open Questions and the Nature of Philosophical Analysis'. In *Themes from G. E. Moore: New Essays in Epistemology and Ethics*, edited by Susana Nuccetelli and Gary Seay, 227–243. Oxford: Oxford University Press, 2007.
- GEACH, PETER. 'Good and Evil'. *Analysis* 17, no. 2 (1956): 33–42.
- GERT, JOSHUA. 'A Functional Role Analysis of Reasons'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 124, no. 3 (2005): 353–378.
- . 'Internalism and Different Kinds of Reasons'. *Philosophical Forum* 34, no. 1 (2003): 53–72.
- . 'Skepticism about Practical Reasons Internalism'. *Southern Journal of Philosophy* 39, no. 1 (2001): 59–77.
- GIBBARD, ALLAN. 'Reasons Thin and Thick'. *Journal of Philosophy* 100, no. 6 (2003): 288–304.
- . *Thinking How to Live*. Cambridge, MA: Harvard University Press, 2003.
- . *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press, 1990.
- GLOCK, HANS-JOHANN. *What Is Analytic Philosophy?* Cambridge: Cambridge University Press, 2008.
- GOLDMAN, ALAN H. 'Desire Based Reasons and Reasons for Desires'. *Southern Journal of Philosophy* 44, no. 3 (Fall 2006): 469–488.
- . 'Reason Internalism'. *Philosophy and Phenomenological Research* 71, no. 3 (2005): 505–532.

- GOLDMAN, ALVIN. 'Internalism Exposed'. In *Epistemology: Internalism and Externalism*, edited by Hilary Kornblith, 207–30. Malden, MA: Blackwell Publishers, 1999/2001.
- GOWER, BARRY. *Scientific Method: An Historical and Philosophical Introduction*. New York: Routledge, 1996.
- HAINES, BYRON. 'Internalism and Moral Training'. *Journal of Value Inquiry* 20, no. 1 (1986): 63–69.
- HAJDIN, MANE. 'External reasons and the foundations of morality: Mother Teresa vs. Thrasymachus'. *Journal of Value Inquiry* 26, no. 3 (1992): 433–441.
- HALDANE, JOHN, and CRISPIN WRIGHT. 'Introduction'. In *Reality, Representation, and Projection*, 3–12. New York: Oxford University Press, 1993.
- HAMILTON, RICHARD PAUL. 'Might There Be Legal Reasons?' *Res Publica: A Journal of Legal and Social Philosophy* 10, no. 4 (2004): 425–447.
- HAMPTON, JEAN E. *The Authority of Reason*. Edited by Richard Healey. UK: Cambridge University Press, 1998.
- HANCOCK, ROGER. 'The Refutation of Naturalism in Moore and Hare'. *Journal of Philosophy* 57 (May 1960): 326–333.
- HARE, R. M. *Freedom and Reason*. Oxford: Clarendon Press, 1963.
- . *The Language of Morals*. London: Oxford University Press, 1952/1967.
- HARMAN, GILBERT. 'Is There a Single True Morality?' In *Morality, Reason and Truth: New Essays on the Foundations of Ethics*, edited by David Copp and David Zimmerman, 27–48. Totowa, NJ: Rowman & Allanheld, 1984/1985.
- . 'Moral Relativism Defended'. *Philosophical Review* 84, no. 1 (1975): 3–22.
- . 'Practical Reasoning'. *Review of Metaphysics* 29 (March 1976): 431–463.
- . *The Nature of Morality: An Introduction to Ethics*. New York: Oxford University Press, 1977.
- . 'What Is Moral Relativism?' In *Values and Morals*, edited by A. I. Goldman and J. Kim. Dordrecht: D. Reidel, 1978.
- HEALTH PROMOTION BOARD, SINGAPORE GOVERNMENT. 'Everyone Has a Reason to Quit Smoking. What's Yours'. *Health Promotion Board*, May 27, 2012. Accessed July 4, 2012. <http://www.lifehacker.com.au/2012/05/reminder-green-tea-contains-caffeine/>.
- HEATHWOOD, CHRIS. 'Moral and Epistemic Open-Question Arguments'. *Philosophical Books* 50, no. 2 (2009): 83–98.
- HERMAN, BARBARA. *The Practice of Moral Judgment*. Cambridge, MA: Harvard University Press, 1993.
- HEWITT, SHARON. 'Normative Qualia and a Robust Moral Realism'. PhD diss., New York University, 2008.
- HILL JR., THOMAS. 'Kantian Normative Ethics'. In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 480–514. Oxford: Oxford University Press, 2006.
- HOOKE, BRAD, and BART STREUMER. 'Procedural and Substantive Practical Rationality'. In *The Oxford Handbook of Rationality*, edited by Alfred R. Mele and Piers Rawling, 57–74. Oxford: Oxford University Press, 2004.
- HORGAN, TERENCE, and MARK TIMMONS. 'Moorean Moral Phenomenology'. In *Themes from G. E. Moore: New Essays in Epistemology and Ethics*, edited by Susana Nuccetelli and Gary Seay, 203–226. Oxford: Oxford University Press, 2007.
- . 'New Wave Moral Realism Meets Moral Twin Earth'. *Journal of Philosophical Research* 16 (1991): 447–465.
- . 'Troubles for Michael Smith's Metaethical Rationalism'. *Philosophical Papers* 25, no. 3 (1996): 203–231.

- . 'Troubles for New Wave Moral Semantics: The Open Question Argument Revived'. *Philosophical Papers* 21, no. 3 (1992): 153–175.
- HUEMER, MICHAEL. *Ethical Intuitionism*. New York: Palgrave Macmillan, 2005.
- HUMBERSTONE, LLOYD. 'Direction of Fit'. *Mind: A Quarterly Review of Philosophy* 101, no. 401 (1992): 59–83.
- HUME, DAVID. *A Treatise of Human Nature*. Edited by Ernest C. Mossner. London: Penguin Books, 1739 & 1740/ 1985.
- HURKA, THOMAS. 'Moore in the Middle'. *Ethics: An International Journal of Social* 113, no. 3 (April 2003): 599–628.
- HURSTHOUSE, ROSALIND. 'Are Virtues the Proper Starting Point for Morality?' In *Contemporary Debates in Moral Theory*, edited by James Dreier, 99–112. Malden, MA: Blackwell Publishing, 2006.
- HURTIG, KENT INGVAR. 'Internalism and Accidie'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 129, no. 3 (2006): 517–543.
- HUTCHESON, FRANCIS. 'A System of Moral Philosophy, Volume I: Concerning the Constitution of Human Nature and the Supreme Good'. In *Four Early Works on Motivation*, edited by Paul McReynolds, 109–280. Gainesville, FL: Scholars' Facsimiles & Reprints, 1755/ 1969.
- JACKSON, FRANK. 'Foreword: Looking Back on the Knowledge Argument'. In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, xv–xix. Cambridge, MA: Bradford Book, MIT Press, 2004.
- . *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon Press, 1998.
- . 'Postscript'. In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 409–15. Cambridge, MA: Bradford Book, MIT Press, 1995/ 2004.
- . 'What Are Cognitivists Doing When They Do Normative Ethics?' *Philosophical Issues: A Supplement to Nous* 15 (2005): 94–106.
- . 'What Mary Didn't Know'. In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 51–6. Cambridge, MA: Bradford Book, MIT Press, 1986/ 2004.
- JACKSON, FRANK, and PHILIP PETTIT. 'Response-dependence without Tears'. *Nous Supplement* 12 (2002): 97–117.
- JACQUETTE, DALE. 'The Blue Banana Trick: Dennett on Jackson's Color Scientist'. *Theoria: A Swedish Journal of Philosophy* 61, no. 3 (1995): 217–230.
- JAMES, WILLIAM. 'Two English Critics'. In *The Meaning of Truth*, chap. 14. New York: Longmans, Green, 1909/ 2009. University of Adelaide Library e-book.
- JENKINS, C. S. 'Realism and Independence'. *American Philosophical Quarterly* 42, no. 3 (July 2005): 199–209.
- . 'What Is Ontological Realism?' *Philosophy Compass* 5, no. 10 (2010): 880–890.
- JOHNSON, ROBERT N. 'Internal Reasons and the Conditional Fallacy'. *Philosophical Quarterly* 49, no. 194 (1999): 53–71.
- . 'Internal Reasons: Reply to Brady, van Roojen and Gert'. *Philosophical Quarterly* 53, no. 213 (2003): 573–80.
- JOHNSTON, MARK. 'Dispositional Theories of Value'. *Aristotelian Society: Supplementary Volume (suppl.)* 63 (1989): 139–174.

- . 'Objectivity Refigured: Pragmatism Without Verificationism'. In *Reality, Representation, and Projection*, 85–130. New York: Oxford University Press, 1993.
- JOYCE, RICHARD. 'Moral Anti-Realism'. Edited by Edward N. Zalta. *Stanford Encyclopedia of Philosophy*, Fall, 2007.
<http://plato.stanford.edu/archives/fall2007/entries/moral-anti-realism/>.
- . 'The Accidental Error Theorist'. In *Oxford Studies in Metaethics vol. 6*, edited by Russ Shafer-Landau, 153–80. Oxford: Oxford University Press, 2011.
- . 'The Error in The Error in the Error Theory'. *Australasian Journal of Philosophy* 89, no. 3 (2011): 519–534.
- . *The Myth of Morality*. Cambridge: Cambridge University Press, 2001. EBL e-book.
- KAGAN, SHELLY. *The Limits of Morality*. New York: Oxford University Press, 1989.
- KAHANE, GUY. 'Feeling Pain for the Very First Time: The Normative Knowledge Argument'. *Philosophy and Phenomenological Research* 80, no. 1 (2010): 20–49.
- . 'Must Metaethical Realism Make a Semantic Claim?' *Journal of Moral Philosophy: An International Journal of Moral* (forthcoming).
- KALDERON, MARK ELI. *Moral Fictionalism*. Oxford: Clarendon Press, 2005.
- KANT, IMMANUEL. *Groundwork for the Metaphysics of Morals*. Edited by Lara Denis, translated by Thomas K. Abbott. Canada: Broadview Press, 1785/ 2005.
- KLAVAN, ANDREW. *The Last Thing I Remember*. Nashville, Tennessee: Thomas Nelson Inc., 2009.
- KNOBE, JOSHUA. 'Intention, Intentional Action and Moral Considerations'. *Analysis* 64, no. 2 (April 2004): 181–187.
- . 'Intentional Action and Side Effects in Ordinary Language'. *Analysis* 63, no. 3 (July 2003): 190–194.
- . 'Intentional Action in Folk Psychology: An Experimental Investigation'. *Philosophical Psychology* 16, no. 2 (2003): 309–324.
- KNYSNA TOURISM. 'Knysna Oyster Festival'. *Shape*, June 2012. Accessed July 4, 2012.
<http://www.shapemag.co.za/event/knysna-oyster-festival/>.
- KORSGAARD, CHRISTINE M. 'Acting for a Reason'. In *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, 207–229. Oxford: Oxford University Press, 2006/ 2008.
- . 'Realism and Constructivism in Twentieth-Century Moral Philosophy'. In *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, 302–326. Oxford: Oxford University Press, 2003/ 2008.
- . *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press, 2009.
- . 'Skepticism about Practical Reason'. *The Journal of Philosophy* 83, no. 1 (1986): 5–25.
- . 'The Normativity of Instrumental Reason'. In *Ethics and Practical Reason*, edited by G. Cullity and B. Gaut, 215–54. Oxford: Oxford University Press, 1997.
- . *The Sources of Normativity*. Edited by Onora O'Neill. Cambridge: Cambridge University Press, 1996.
- KOSSO, PETER. *A Summary of Scientific Method*. Springer, 2011.
- LAWRENCE, GAVIN. 'The Rationality of Morality'. In *Virtues and Reasons: Philippa Foot and Moral Theory*, edited by Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn, 89–147. New York: Clarendon Press, 1995.

- LEBAR, MARK. 'Three Dogmas of Response-Dependence'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 123, no. 3 (April 2005): 175–211.
- LENMAN, JAMES. 'Belief, Desire and Motivation: An Essay in Quasi-Hydraulics'. *American Philosophical Quarterly* 33, no. 3 (July 1996): 291–301.
- . 'Humean Constructivism in Moral Theory'. In *Oxford Studies in Metaethics vol. 5*, edited by Russ Shafer-Landau, 175–193. Oxford: Oxford University Press, 2010.
- . 'Reasons for Action: Justification vs. Explanation'. Edited by Edward N. Zalta. *Stanford Encyclopedia of Philosophy*, Spring, 2009.
<http://plato.stanford.edu/archives/spr2009/entries/reasons-just-vs-expl/>
- LEWIS, DAVID. 'Dispositional Theories of Value'. *Aristotelian Society: Supplementary Volume* 63 (1989): 113–137.
- . 'Radical Interpretation'. *Synthese: An International Journal for Epistemology* 27 (August 1974): 331–344.
- . 'What Experience Teaches'. In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 77–103. Cambridge, MA: Bradford Book, MIT Press, 1988/ 2004.
- LILLEHAMMER, HALLVARD. 'The Doctrine of Internal Reasons'. *Journal of Value Inquiry* 34, no. 4 (2000): 507–516.
- LOAR, BRIAN. 'Phenomenal States (Revised)'. In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 219–39. Cambridge, MA: Bradford Book, MIT Press, 1997/ 2004.
- LOCKE, DON. 'Reasons, Wants, and Causes'. *American Philosophical Quarterly* 11, no. 3 (1974): 169–179.
- LOUX, MICHAEL J. *Metaphysics: A Contemporary Introduction*. 3rd ed. New York: Routledge, 2006.
- LYCAN, WILLIAM G. *Consciousness and Experience*. MIT: MIT Press, 1996.
- MACKIE, J. L. *Ethics: Inventing Right and Wrong*. London: Penguin Books, 1977.
- MARKOVITS, JULIA. 'Internal Reasons and the Motivating Intuition'. In *New Waves in Metaethics*, edited by Michael S. Brady, 141–65. Basingstoke: Palgrave Macmillan, 2010.
- . 'Why Be an Internalist About Reasons?' In *Oxford Studies in Metaethics vol. 6*, edited by Russ Shafer-Landau, 255–79. Oxford: Oxford University Press, 2011.
- MCCLOSKEY, H. J. 'An Examination of Restricted Utilitarianism'. *Philosophical Review* 66, no. 4 (1957): 466–485.
- MCDOWELL, JOHN. 'Are Moral Requirements Hypothetical Imperatives?' In *Mind, Value, & Reality*, edited by John McDowell, 77–94. Cambridge, MA: Harvard University Press, 1978/ 1998.
- . 'Might There Be External Reasons?' In *World, Mind, and Ethics*, edited by J. E. J. Altham, 68–85. New York: Cambridge University Press, 1995.
- . *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press, 1998.
- . 'The Role of Eudaimonia in Aristotle's Ethics'. In *Mind, Value, & Reality*, 3–22. Cambridge, MA: Harvard University Press, 1980/ 1998.
- . 'Values and Secondary Qualities'. In *Morality and Objectivity*, edited by Ted Honderich, 110–129. London: Routledge Kegan & Paul, 1985.
- . 'Virtue and Reason'. In *Mind, Value, & Reality*, edited by John McDowell, 50–73. Cambridge, MA: Harvard University Press, 1979/ 1998.

- MCNAUGHTON, DAVID. *Moral Vision: An Introduction to Ethics*. Oxford: Basil Blackwell, 1988.
- MCNAUGHTON, DAVID, and PIERS RAWLING. 'Naturalism and Normativity'. *Aristotelian Society: Supplementary Volume* Supp, no. 77 (2003): 23–45.
- . 'Value and Agent-Relative Reasons'. *Utilitas: A Journal of Utilitarian Studies* 7, no. 1 (May 1995): 31–47.
- MELE, ALFRED R. 'Action'. In *The Oxford Handbook of Contemporary Philosophy*, edited by Frank Jackson and Michael Smith, 334–357. Oxford: Oxford University Press, 2005.
- . *Motivation and Agency*. New York: Oxford University Press, 2003.
- . *Springs of Action: Understanding Intentional Behavior*. Oxford: Oxford University Press, 1992.
- MILLER, ALEXANDER. *An Introduction to Contemporary Metaethics*. Cambridge: Polity Press, 2003.
- . 'Realism'. Edited by Edward N. Zalta. *Stanford Encyclopedia of Philosophy*, Summer, 2010.
<http://plato.stanford.edu/archives/sum2010/entries/realism/>.
- MILLGRAM, ELIJAH. 'Was Hume a Humean?' *Hume Studies* 21, no. 1 (April 1995): 75–93.
- . 'Williams' Argument Against External Reasons'. *Noûs* 30, no. 2 (1996): 197–220.
- MILLIKEN, JOHN. 'In a Fitter Direction: Moving Beyond the Direction of Fit Picture of Belief and Desire'. *Ethical Theory and Moral Practice: An International Forum* 11, no. 5 (2008): 563–571.
- MILO, RONALD D. *Immorality*. Princeton, NJ: Princeton University Press, 1984.
- MOORE, ANDREW. 'Time and Well-Being'. In *Time and Ethics: Essays at the Intersection*, edited by Heather Dyke, 85–97. Dordrecht: Kluwer Academic Publishers, 2003.
- MOORE, G. E. 'G.E. Moore: A Reply to My Critics'. In *The Philosophy of G.E. Moore*, edited by P. A. Schilpp. La Salle, IL: Open Court, 1942.
- . 'Preface to the Second Edition'. In *Principia Ethica*, edited by Thomas Baldwin. Cambridge: Cambridge University Press, unknown/ 1993.
- . *Principia Ethica*. Edited by Thomas Baldwin. 2nd ed. Cambridge: Cambridge University Press, 1903/ 1993.
- NAGEL, THOMAS. *The Last Word*. New York: Oxford University Press, 1997.
- . *The Possibility of Altruism*. Oxford: Clarendon Press, 1970.
- . *The View From Nowhere*. New York: Oxford University Press, 1986.
- NEMIROW, LAURENCE. 'Mortal Questions. by Thomas Nagel'. *The Philosophical Review* 89, no. 3 (July 1, 1980): 473–477.
- NIDA-RÜMELIN, MARTINE. 'What Mary Couldn't Know: Belief About Phenomenal States'. In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 241–67. Cambridge, MA: The MIT Press, 1995/ 2004.
- NOLAN, DANIEL. 'The Extent of Metaphysical Necessity'. *Nous-Supplement: Philosophical Perspectives* 25, Metaphysics, no. 1 (2011): 313–39.
- NOTTELMANN, NIKOLAJ. 'Belief-Desire Explanation'. *Philosophy Compass* 6, no. 12 (December 1, 2011): 912–921.
- NUCCETELLI, SUSANA, and GARY SEAY. 'What's Right with the Open Question Argument'. In *Themes from G. E. Moore: New Essays in Epistemology and Ethics*,

- edited by Susana Nuccetelli and Gary Seay, 261-282. Oxford: Oxford University Press, 2007.
- ODDIE, GRAHAM. *Value, Reality, and Desire*. Oxford: Oxford University Press, 2005.
- OWENS, DAVID. 'Rationalism about Obligation'. *European Journal of Philosophy* 16, no. 3 (December 2008): 403-431.
- PARFIT, DEREK. *On What Matters: Volume Two*. Oxford: Oxford University Press, 2011.
- . 'Rationality and Reasons'. In *Exploring Practical Philosophy: From Action to Values*, edited by Dan Egonsson, Jonas Josefsson, Bjorn Peterson, and Toni Rønnow-Rasmussen, 17-39. Aldershot: Ashgate Publishing, 2001.
- . 'Reasons and Motivation'. *Proceedings of the Aristotelian Society Supplementary Volume* 71 (1997): 99-130.
- . *Reasons and Persons*. Oxford: Clarendon Press, 1984.
- PEACOCKE, CHRISTOPHER. 'Moral Rationalism'. *Journal of Philosophy* 101, no. 10 (October 2004): 499-526.
- PETTIT, PHILIP. 'Motion Blindness and the Knowledge Argument'. In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 105-42. Cambridge, MA: Bradford Book, MIT Press, 2004.
- . 'Realism and Response-Dependence'. *Mind: A Quarterly Review of Philosophy* 100, no. 400 (1991): 587-626.
- . 'Universalizability without Utilitarianism'. *Mind: A Quarterly Review of Philosophy* 96, no. 381 (1987): 74-82.
- PETTIT, PHILIP, and MICHAEL SMITH. 'Backgrounding Desire'. *Philosophical Review* (October 1990): 565-592.
- . 'External Reasons'. In *McDowell and His Critics*, edited by Cynthia Macdonald and Graham Macdonald, 142-179. Malden, MA: Blackwell Publishing, 2006.
- PIGDEN, CHARLES. 'Desiring to Desire: Russell, Lewis, and G. E. Moore'. In *Themes from G. E. Moore: New Essays in Epistemology and Ethics*, edited by Susana Nuccetelli and Gary Seay, 244-60. Oxford: Oxford University Press, 2007.
- PLATO. 'Meno'. In *Readings in Ancient Greek Philosophy: From Thales to Aristotle*, edited by S. Marc Cohen, Patricia Curd, and C. D. C. Reeve. 2nd ed. Indianapolis: Hackett Publishing Company, 347 B.C./ 2000.
- . *The Republic*. London: Penguin Classics, 347AD/ 2007.
- POLLOCK, JOHN, and JOSEPH CRUZ. *Contemporary Theories of Knowledge*. 2nd ed. Maryland: Rowman & Littlefield Publishers, Inc., 1999.
- PORTMORE, DOUGLAS W. 'McNaughton and Rawling on the Agent-Relative/Agent-Neutral Distinction'. *Utilitas: A Journal of Utilitarian Studies* 13, no. 3 (2001): 350-356.
- PRICHARD, H. A. 'Acting, Willing, Desiring'. In *Moral Obligation and Duty and Interest: Essays and Lectures by H. A. Prichard*, 187-98. Oxford: Oxford University Press, 1945/ 1968.
- . 'Does Moral Philosophy Rest on a Mistake?' In *Moral Obligation and Duty and Interest: Essays and Lectures by H. A. Prichard*, 1-17. Oxford: Oxford University Press, 1912/ 1968.
- . 'Duty and Interest'. In *Moral Obligation and Duty and Interest: Essays and Lectures by H. A. Prichard*, 201-238. Oxford: Oxford University Press, 1928/ 1968.
- PUTNAM, HILARY. *Reason, Truth and History*. Cambridge: Cambridge University Press, 1981.

- . *The Many Faces of Realism: The Paul Carus Lectures*. La Salle, IL: Open Court Publishing Company, 1987.
- QUINN, WARREN. 'Putting Rationality in Its Place'. In *Virtues and Reasons: Philippa Foot and Moral Theory*, edited by Rosalind Hursthouse, 181–208. New York: Clarendon Press, 1993/ 1995.
- RAILTON, PETER. 'Facts and Values'. *Philosophical Topics* 14, no. 2 (2003 1986): 5–31.
- . 'Humean Theory of Practical Rationality'. In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 265–81. New York: Oxford University Press, 2006.
- . 'Moral Realism'. *Philosophical Review* 95, no. 2 (1986): 163–207.
- . 'Naturalism and Prescriptivity'. *Social Philosophy and Policy* 7, no. 1 (1989): 151–174.
- . 'Normative Force and Normative Freedom: Hume and Kant, but not Hume versus Kant'. *Ratio (New Series)* 12, no. 4 (1999): 320–353.
- . 'Red, Bitter, Good'. In *European Review of Philosophy, Volume 3: Response-Dependence*, edited by Roberto Casati and Christine Tappolet, 67–84. Stanford: CSLI Publications, 1998.
- RAZ, JOSEPH. 'Agency, Reason, and the Good'. In *Engaging Reason: On the Theory of Value and Action*, 22–45. Oxford: Oxford University Press, 1999.
- . *Practical Reason and Norms*. Princeton: Princeton University Press, 1975.
- . 'Reasons: Normative and Explanatory'. In *New Essays on the Explanation of Action*, edited by C. Sandis, 184–202. Basingstoke: Palgrave, 2009.
- REGAN, DONALD. 'How to Be a Moorean'. *Ethics: An International Journal of Social* 113, no. 3 (2003): 651–77.
- RIDGE, MICHAEL. 'Moral Non-Naturalism'. Edited by Edward N. Zalta. *Stanford Encyclopedia of Philosophy*, Fall, 2008.
<http://plato.stanford.edu/archives/fall2008/entries/moral-non-naturalism/>.
- . 'Reasons for Action: Agent-Neutral vs. Agent-Relative'. Edited by Edward N. Zalta. *Stanford Encyclopedia of Philosophy*, Fall, 2008.
<http://plato.stanford.edu/archives/fall2008/entries/reasons-agent/>
- ROBB, ANDREW. 'Mornings the Darkest Time of Day'. *The Australian*, August 27, 2011. Accessed August 30, 2011.
<http://www.theaustralian.com.au/national-affairs/health-of-the-nation/mornings-the-darkest-time-of-day/story-fn9iqmqf-1226122668395>.
- ROBBINS, LIZ. 'Supporting Cast'. *Runner's World*, 2008.
- ROBERTSON, SIMON. 'Introduction: Normativity, Reasons, Rationality'. In *Spheres of Reason: New Essays in the Philosophy of Normativity*, 1–28. Oxford: Oxford University Press, 2009.
- ROBERTSON, TERESA. 'Internalism, (Super)fragile Reasons, and the Conditional Fallacy'. *Philosophical Papers* 32, no. 2 (July 2003): 171–184.
- ROBINSON, HOWARD. 'Dennett on the Knowledge Argument'. In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 69–73. Cambridge, MA: Bradford Book, MIT Press, 1993/ 2004.
- RØNNOW-RASMUSSEN, TONI. 'Normative Reasons and the Agent-Neutral/Relative Dichotomy'. *Philosophia: Philosophical Quarterly of Israel* 37, no. 2 (2009): 227–243.
- ROSATI, CONNIE S. 'Agency and the Open Question Argument'. *Ethics: An International Journal of Social* 113, no. 3 (April 2003): 490–527.
- . 'Moral Motivation'. Edited by Edward N. Zalta. *Stanford Encyclopedia of Philosophy*, Fall, 2008.

- <http://plato.stanford.edu/archives/fall2008/entries/moral-motivation/>
- . 'Naturalism, Normativity, and the Open Question Argument'. *Noûs* 29, no. 1 (March 1995): 46–70.
- . 'Persons, Perspectives, and Full Information Accounts of the Good'. *Ethics: An International Journal of Social* 105, no. 2 (1995): 296–325.
- ROSS, JACOB. 'How to Be a Cognitivist about Practical Reason'. In *Oxford Studies in Metaethics vol. 4*, edited by Russ Shafer-Landau, 243–282. Oxford: Oxford University Press, 2009.
- ROSS, W. D. *The Right and the Good*. Oxford: Clarendon Press, 1930/ 1967.
- RUSSELL, BERTRAND. 'William James's Conception of Truth'. In *Philosophical Essays*, 104–122. London: Routledge, 1908/ 2009. EBL e-book.
- RYAN, SHARON. 'Doxastic Compatibilism and the Ethics of Belief'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 114, no. 1–2 (May 2003): 47–79.
- RYLE, GILBERT. *The Concept of Mind*. Chicago: University of Chicago Press, 1949/ 2000.
- SAYRE-MCCORD, GEOFFREY. 'Being a Realist about Relativism (In Ethics)'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 61, no. 1/2 (1991): 155–176.
- . 'Moral Realism [Oxford Handbook]'. In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 39–62. New York: Oxford University Press, 2006.
- . 'Moral Realism [SEP]'. Edited by Edward N. Zalta. *Stanford Encyclopedia of Philosophy*, Summer, 2011.
<http://plato.stanford.edu/archives/sum2011/entries/moral-realism/>
- . 'The Many Moral Realisms'. *Southern Journal of Philosophy (Suppl.)* 24 (1986): 1–22.
- SCANLON, T. M. 'Metaphysics and Morals'. *Proceedings of the American Philosophical Association* 77, no. 2 (2003): 7–22.
- . *What We Owe to Each Other*. Cambridge, MA: Belknap Press of Harvard University Press, 1998.
- SCHEMBRI, JIM. 'The Academy Awards Rebooted'. *The Age*, March 5, 2010. Accessed August 30, 2011. <http://www.theage.com.au/entertainment/oscars/the-academy-awards-rebooted-20100305-pm11.html>.
- SCHROEDER, MARK. 'Buck-Passers' Negative Thesis'. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action* 12, no. 3 (September 2009): 341–347.
- . 'Realism and Reduction: The Quest for Robustness'. *Philosophers' Imprint* 5, no. 1 (February 2005): 1–18.
- . 'Reasons and Agent-Neutrality'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 135, no. 2 (September 2007): 279–306.
- . 'Reply to Shafer-Landau, Mcpherson, and Dancy'. *Philosophical Studies* 157, no. 3 (2012): 463–474.
- . *Slaves of the Passions*. Oxford: Oxford University Press, 2007.
- . 'Teleology, Agent-Relative Value, and "Good"'. *Ethics: An International Journal of Social* 117, no. 2 (2007): 265–295.
- . 'The Humean Theory of Reasons'. In *Oxford Studies in Metaethics vol. 2*, edited by Russ Shafer-Landau, 195–219. Oxford: Oxford University Press, 2007.
- SCHROEDER, TIM. 'Desire'. Edited by Edward N. Zalta. *Stanford Encyclopedia of Philosophy*, Winter, 2009.
<http://plato.stanford.edu/archives/win2009/entries/desire/>

- SCHROETER, FRANÇOIS. 'Reflective Equilibrium and Antitheory'. *Noûs* 38, no. 1 (March 2004): 110–134.
- SCHUELER, G. F. *Desire: Its Role in Practical Reason and the Explanation of Action*. Cambridge, MA: The MIT Press, 1995.
- . 'Pro-Attitudes and Direction of Fit'. *Mind: A Quarterly Review of Philosophy* 100, no. 398 (April 1991): 277–281.
- . *Reasons and Purposes: Human Rationality and the Teleological Explanation of Action*. Oxford: Clarendon Press, 2003.
- SETIYA, KIERAN. 'Cognitivism about Instrumental Reason'. *Ethics: An International Journal of Social* 117, no. 4 (July 2007): 649–673.
- . 'Practical Knowledge'. *Ethics: An International Journal of Social* 118, no. 3 (April 2008): 388–409.
- . 'Practical Knowledge Revisited'. *Ethics* 120, no. 1 (October 1, 2009): 128–137.
- . *Reasons without Rationalism*. Princeton, NJ: Princeton University Press, 2007.
- SHAFFER-LANDAU, RUSS. *Moral Realism: A Defence*. Oxford: Clarendon Press, 2003.
- . 'Replies to Critics'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 126, no. 2 (2005): 313–329.
- . 'Three Problems for Schroeder's Hypotheticalism'. *Philosophical Studies* 157, no. 3 (2012): 435–443.
- SHOEMAKER, DAVID. 'Ethics Discussions at PEA Soup: Mark van Roojen's "Moral Rationalism and Rational Amoralism," with Commentary by Russ Shafer-Landau'. *PEA Soup*, July 2010. Accessed August 5, 2010.
<http://peasoup.typepad.com/peasoup/2010/07/ethics-discussions-at-pea-soup-mark-van-roojens-moral-rationalism-and-rational-amoralism-with-commen.html>.
- SHOEMAKER, SYDNEY. 'Self-Knowledge and "Inner Sense": Lecture III: The Phenomenal Character of Experience'. *Philosophy and Phenomenological Research* 54, no. 2 (June 1, 1994): 291–314.
- SHOPE, ROBERT K. 'The Conditional Fallacy in Contemporary Philosophy'. *Journal of Philosophy* 75, no. 8 (1978): 397–413.
- SIDGWICK, HENRY. *Methods of Ethics*. 7th ed. Cambridge: Hackett Publishing Company, 1907.
- SINGER, PETER. 'Famine, Affluence, and Morality'. *Philosophy and Public Affairs* 1, no. 1972 (1972): 229–243.
- . *The Life You Can Save: Acting Now to End World Poverty*. Melbourne: Text Publishing, 2009.
- SINHABABU, NEIL. 'The Humean Theory of Motivation Reformulated and Defended'. *The Philosophical Review* 118, no. 4 (2009): 465–500.
- SKORUPSKI, JOHN. 'Agent-Neutrality, Consequentialism, Utilitarianism... A Terminological Note'. *Utilitas: A Journal of Utilitarian Studies* 7, no. 1 (May 1995): 49–54.
- . 'Internal Reasons and the Scope of Blame'. In *Bernard Williams*, edited by Alan Thomas, 71–103. Cambridge: Cambridge University Press, 2007.
- . 'Irrealist Cognitivism'. *Ratio: An International Journal of Analytic Philosophy* 12, no. 4 (December 1999): 436–459.
- . 'Neutral versus Relative: A Reply to Broome, and McNaughton and Rawling'. *Utilitas: A Journal of Utilitarian Studies* 8, no. 2 (July 1996): 235–248.
- . *The Domain of Reasons*. Oxford: Oxford University Press, 2010. Oxford Scholarship Online.
- . 'The Ontology of Reasons'. *Topoi: An International Review of Philosophy* 21, no. 1–2 (2002): 113–124.

- . 'The Unity and Diversity of Reasons'. In *Spheres of Reason: New Essays in the Philosophy of Normativity*, edited by Simon Robertson, 113–39. Oxford: Oxford University Press, 2009.
- SLOTE, MICHAEL. *Morals from Motives*. Oxford: Oxford University Press, 2001.
- SMITH, MICHAEL. 'Dispositional Theories of Value'. *Aristotelian Society: Supplementary Volume (suppl.)* 63 (1989): 89–111.
- . 'Humean Rationality'. In *The Oxford Handbook of Rationality*, 75–92. Oxford: Oxford University Press, 2004. Oxford Scholarship Online.
- . 'In Defence of The Moral Problem: A Reply to Brink, Copp, and Sayre-McCord'. In *Ethics and the A Priori*, 259–96. Cambridge: Cambridge University Press, 1997/2004. EBL e-book.
- . 'Internal Reasons'. *Philosophy and Phenomenological Research* 55, no. 1 (1995): 109–131.
- . 'Moral Realism'. In *Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-ethics*, 181–207. Cambridge: Cambridge University Press, 2000/2004. EBL e-book.
- . 'Objectivity and Moral Realism: On the Significance of the Phenomenology of Moral Experience'. In *Reality, Representation, and Projection*, 235–255. New York: Oxford University Press, 1993.
- . 'The Humean Theory of Motivation'. *Mind: A Quarterly Review of Philosophy* 96, no. 381 (1987): 36–61.
- . *The Moral Problem*. Malden, MA: Blackwell Publishing, 1994.
- SMITH, MICHAEL, and DANIEL STOLJAR. 'Global Response-Dependence and Noumenal Realism'. *Monist: An International Quarterly Journal of General Philosophical Inquiry* 81, no. 1 (1998): 85–111.
- SNARE, FRANK. 'The Open Question as a Linguistic Test'. In *Arguing About Metaethics*, edited by Andrew Fisher and Simon Kirchin, 59–65. London: Routledge, 1975/2006.
- SOBEL, DAVID. 'Explanation, Internalism, and Reasons for Action'. *Social Philosophy and Policy* 18, no. 2 (Summer 2001): 218–235.
- . 'Full Information Accounts of Well-Being'. *Ethics: An International Journal of Social* 104, no. 4 (July 1994): 784–810.
- . 'Reply to Robertson'. *Philosophical Papers* 32, no. 2 (July 2003): 185–191.
- SOBEL, DAVID, and DAVID COPP. 'Against Direction of Fit Accounts of Belief and Desire'. *Analysis* 61, no. 1 (2001): 44–53.
- SOCRATES. 'Phaedo'. In *Readings in Ancient Greek Philosophy: From Thales to Aristotle*, edited by S. Marc Cohen, Patricia Curd, and C. D. C. Reeve. 2nd ed. Indianapolis: Hackett Publishing Company, 381 B.C./2000.
- STAMPE, DENNIS W. 'The Authority of Desire'. *Philosophical Review* 96 (1987): 335–81.
- STRANDBERG, CAJ. 'In Defence of the Open Question Argument'. *Journal of Ethics: An International Philosophical Review* 8, no. 2 (2004): 179–196.
- STRATTON-LAKE, PHILIP. 'Introduction'. In *Ethical Intuitionism: Re-evaluations*. Oxford: Clarendon Press, 2002.
- STRATTON-LAKE, PHILIP, and BRAD HOOKER. 'Scanlon versus Moore on Goodness'. In *Metaethics after Moore*, edited by Terry Horgan and Mark Timmons, 149–168. Oxford: Clarendon Press, 2006.
- STREET, SHARON. 'A Darwinian Dilemma for Realist Theories of Value'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 127, no. 1 (2006): 109–166.
- . 'Constructivism about Reasons'. In *Oxford Studies in Metaethics vol. 3*, edited by Russ Shafer-Landau, 207–45. Oxford: Oxford University Press, 2008.

- . 'Mind-Independence Without the Mystery: Why Quasi-Realists Can't Have It Both Ways'. In *Oxford Studies in Metaethics vol. 6*, edited by Russ Shafer-Landau, 1–32. Oxford: Oxford University Press, 2011.
- . 'What Is Constructivism in Ethics and Metaethics?' *Philosophy Compass* 5, no. 5 (May 2010): 363–384.
- STURGEON, NICHOLAS. 'Moore on Ethical Naturalism'. *Ethics* 113 (2003): 528–56.
- . 'What Difference Does It Make Whether Moral Realism Is True?' *The Southern Journal of Philosophy, Supplement* 24 (1986): 115–141.
- TÄNNISJÖ, TORBJÖRN. *Moral Realism*. Maryland: Rowman & Littlefield Publishers, Inc., 1990.
- TENENBAUM, SERGIO. 'Direction of Fit and Motivational Cognitivism'. In *Oxford Studies in Metaethics vol. 1*, edited by Russ Shafer-Landau. Oxford: Oxford University Press, 2006.
- . 'Ethical Internalism and Glaucon's Question'. *Noûs* 34, no. 1 (March 2000): 108–130.
- . 'The Conclusion of Practical Reason'. In *Moral Psychology*, edited by Sergio Tenenbaum, 323–343. New York: Rodopi, 2007.
- THOMPSON, BRAD. 'Moral Value, Response-Dependence, and Rigid Designation'. *Canadian Journal of Philosophy* 36, no. 1 (March 2006): 71–94.
- THOMSON, JUDITH JARVIS. 'Normativity'. In *Oxford Studies in Metaethics vol. 2*, edited by Russ Shafer-Landau, 240–265. Oxford: Oxford University Press, 2007.
- TYE, MICHAEL. 'Knowing What It Is Like: The Ability Hypothesis and the Knowledge Argument'. In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 143–160. Cambridge, MA: Bradford Book, The MIT Press, 2000/2004.
- UNGER, PETER K. *Living High and Letting Die: Our Illusion of Innocence*. New York: Oxford University Press, 1996. Oxford Scholarship Online.
- VAN ROOJEN, MARK. 'Moral Rationalism and Rational Amoralism'. *Ethics: An International Journal of Social* 120, no. 3 (2010): 495–525.
- . 'Rationalist Realism and Constructivist Accounts of Morality'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 126, no. 2 (2005): 285–295.
- VELLEMAN, J. DAVID. *How We Get Along*. Cambridge: Cambridge University Press, 2009.
- . 'Introduction'. In *The Possibility of Practical Reason*, 1–31. Oxford: Clarendon Press, 2000.
- . 'On the Aim of Belief'. In *The Possibility of Practical Reason*, 244–81. Oxford: Clarendon Press, 2000.
- . 'The Guise of the Good'. In *The Possibility of Practical Reason*, 99–122. Oxford: Clarendon Press, 1992/2000.
- . 'The Possibility of Practical Reason'. In *The Possibility of Practical Reason*, 170–99. Oxford: Clarendon Press, 1996/2000.
- . 'The Possibility of Practical Reason'. *Ethics: An International Journal of Social* 106, no. 4 (July 1996): 694–726.
- . *The Possibility of Practical Reason*. Oxford: Clarendon Press, 2000.
- . 'What Happens When Someone Acts?' In *The Possibility of Practical Reason*, 123–43. Oxford: Clarendon Press, 1992/2000.
- VOGLER, CANDACE. *Reasonably Vicious*. Cambridge, MA: Harvard University Press, 2002.

- WALLACE, R. JAY. 'How to Argue about Practical Reason'. *Mind: A Quarterly Review of Philosophy* 99, no. 395 (July 1990): 355–385.
- . 'Moral Psychology'. In *The Oxford Handbook of Contemporary Philosophy*, edited by Frank Jackson and Michael Smith, 86–113. Oxford: Oxford University Press, 2005.
- WALTERNEWTON. 'Best Program to Play WAV Files In?' *iBreatheMusic*, May 5, 2011. Accessed August 30, 2011. <http://www.ibreathemusic.com/forums/showthread.php?p=147232>.
- WEDGWOOD, RALPH. 'Choosing Rationally and Choosing Correctly'. In *Weakness of Will and Practical Irrationality*, edited by Sarah Stroud, 201–229. Oxford: Clarendon Press, 2003.
- . 'Practical Reason and Desire'. *Australasian Journal of Philosophy* 80, no. 3 (September 2002): 345–358.
- . 'The Essence of Response-Dependence'. In *European Review of Philosophy, Volume 3: Response-Dependence*, edited by Roberto Casati and Christine Tappolet, 31–54. Stanford: CSLI Publications, 1998.
- . *The Nature of Normativity*. Oxford: Clarendon Press, 2007.
- WHITE, NICHOLAS P. 'Inquiry'. *Review of Metaphysics* 28 (December 1974): 289–310.
- WIGGINS, DAVID. 'A Neglected Position?' In *Reality, Representation, and Projection*, edited by Crispin Wright and John Haldane. New York: Oxford University Press, 1993.
- . 'A Sensible Subjectivism?' In *Needs, Values, Truth*, 185–211. Oxford: Basil Blackwell, 1987.
- . 'Cognitivism, Naturalism, and Normativity: A Reply to Peter Railton'. In *Reality, Representation, and Projection*, 301–313. New York: Oxford University Press, 1993.
- . 'Moral Cognitivism, Moral Relativism and Motivating Moral Beliefs'. *Proceedings of the Aristotelian Society* 91 (1991): 61–85.
- . 'Objectivity in Ethics; Two Difficulties, Two Responses'. *Ratio: An International Journal of Analytic Philosophy* 18, no. 1 (March 2005): 1–26.
- . 'Truth, Invention and the Meaning of Life'. *Proceedings of the British Academy* 62 (1976): 331–78.
- WILLIAMS, BERNARD. *Ethics and the Limits of Philosophy*. 1st Edition, 3rd Impression with amendments. London: Fontana Press, 1985.
- . *Ethics and the Limits of Philosophy*. London: Fontana Press, 1985.
- . 'Internal and External Reasons'. In *Moral Luck*, edited by Bernard Williams, 101–113. New York: Cambridge University Press, 1979/ 1981.
- . 'Internal Reasons and the Obscurity of Blame'. In *Making Sense of Humanity: And Other Philosophical Papers*, 35–45. Great Britain: Cambridge University Press, 1989/ 1995.
- . 'Persons, Character and Morality'. In *Moral Luck: Philosophical Papers 1973–1980*, edited by Bernard Williams, 1–19. Cambridge: Cambridge University Press, 1981.
- . 'Postscript: Some Further Notes on Internal and External Reasons'. In *Varieties of Practical Reasoning*, edited by Elijah Millgram, 91–7. Cambridge, MA: MIT Press, 2001.
- . 'Replies'. In *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, edited by J. E. J. Altham and Ross Harrison, 185–224. New York: Cambridge University Press, 1995.
- WISDOM, JEFFREY ROBERT. 'From Moral Fictionalism to Moral Realism: An Essay in Moral Metaphysics'. PhD diss., University of Connecticut, 2010.

- WONG, DAVID B. 'Moral Reasons: Internal and External'. *Philosophy and Phenomenological Research* 72, no. 3 (2006): 536–558.
- WRIGHT, CRISPIN. 'Moral Values, Projection and Secondary Qualities'. *Aristotelian Society: Supplementary Volume (suppl.)* 62 (1988): 1–26.
- . *Truth and Objectivity*. Cambridge, MA: Harvard University Press, 1992.
- ZARPENTINE, CHRIS. 'Michael Smith, Rationalism, and the Moral Psychology of Psychopathy'. *Florida Philosophical Review: The Journal of the Florida Philosophical Association* 7, no. 1 (Summer 2007): 1–15.
- ZIMMERMAN, AARON. *Moral Epistemology*. London: Routledge, 2010.

Index

- agent-neutral 4, 9, 10, 14, 16-26, 30, 78, 88, 92-95, 97, 98, 109-113, 115-117, 119-225, 231, 282
 - explanatory reason 92
 - justifying reason 25, 110
 - reason 15-17, 22, 25
- agent-relational 16, 17, 20-26, 88, 92, 93, 112, 113, 116, 123
 - explanatory reason 91
 - justifying reason 17
 - justifying reason* 23
 - reason 16, 17, 22, 25
- agent-relative 14-19, 21-26, 78, 79, 88, 90-95, 97, 98, 109-112, 114, 115, 121, 224, 225, 245
 - explanatory 88
 - explanatory reason 92
 - justifying reason 24
 - reason 15, 19, 20
- anti-realism 11, 251, 291
- Blackburn, Simon 27, 43, 75, 76, 97, 101, 105-107, 191, 199, 214, 247, 253, 254, 260, 262, 267, 271, 278, 279, 284
- Brink David O. 1, 2, 9, 11, 44, 54, 67, 72, 200, 201, 233
- cognitivism 2, 6, 8, 27, 126
- conclusive reason 35, 137, 152, 155, 172, 173, 197, 225, 229, 231, 232
- Cuneo, Terence 6, 8, 11, 26, 28-30
- Dancy, Jonathan 3, 9, 12-14, 21-23, 31, 33-35, 37, 81, 83, 92, 93, 96, 97, 143, 158, 233, 234, 286, 296
- Darwall, Stephen 43, 75, 76, 81, 98, 133, 229
- dependence 9, 10, 14, 95, 245, 246, 248, 255, 257, 282, 290
 - essential 14, 258
- desire 18, 19, 44, 80-83, 92, 94, 98, 100, 102, 125, 128, 131, 148, 189, 268, 296
- EA *See* evaluative attitude
- evaluative attitude 84, 103, 125, 128, 134, 239
 - EA 81, 84-87, 91-94, 97, 98, 100-103, 105, 109, 115
- explanatory reason 32, 33, 36, 37, 79, 88-96, 109, 110, 115, 157, 200, 225
 - objective 37
 - subjective 37
- extension-determining 249, 251, 254, 272, 273, 281
- extension-reflecting 249, 251-254, 271, 272
- external reason 127, 141-143, 152, 156, 158, 160, 177, 178, 180, 187, 190, 208, 212, 214, 218, 219, 220, 222
 - External Reason* 141, 143, 152
 - external-satisfying 142, 143
 - internal *See* internal reason
 - pure 142, 143, 149, 155-157, 159, 160, 169, 171, 172, 186, 187, 190, 196, 199, 200, 201, 207, 218-220, 222, 224, 275
- external reason statement 128, 140, 141, 155, 158-160, 163-165, 168-171, 177, 191, 215
- External Reason* *See* external reason
- Finlay, Stephen 9, 27-29, 37, 83, 94, 97, 125, 128, 135, 143, 145, 153, 158, 161, 198, 199, 223, 251, 287, 288
- Harman, Gilbert 27, 28, 30, 74, 80, 107, 158, 170, 182, 199, 215, 238, 289
- HTM *See* The Humean Theory of Motivation
- Huemer, Michael 9, 27, 42, 54, 82, 245
- ideal observer 26, 38, 148, 251, 252, 273-277
- idealism 10
- Immodest Externalism 152, 155, 159, 160, 173, 181, 196-198, 203, 206, 275

- impossible 12, 87, 111-113, 130, 138, 166, 174, 176, 182-184, 191, 207, 209-211, 275, 280
 - deterministically 111, 182-184
 - logically 175, 176, 182, 275
 - physically 182, 183, 211
 - rationally 111, 191
- independence 2, 6, 9, 10, 12-14, 143, 186, 255
 - causal 11
 - essential 12-14, 186, 248, 255
 - modal 12, 13, 248, 255, 256, 258, 272
- internal reason 128-130, 140-144, 153, 159, 179, 180, 198, 199, 201, 203, 204, 210, 212, 216, 218, 219, 220, 275
- internal reason statement 127, 128, 141, 160, 177, 199, 215
- Jackson, Frank 49-51, 54, 55, 58, 60, 246, 248, 257, 285, 287, 290, 292, 293, 294, 295, 299, 300
- Jenkins, C. S. 3, 9, 11-13
- Johnson, Robert N. 146-150, 155, 198
- Joyce, Richard 2, 9, 11, 27, 33, 59, 77, 93, 148, 158, 169, 198, 199, 209, 231, 233, 240, 251, 255, 257, 274, 287, 291
- Judicious Agent Theory 38, 39, 246, 273, 281-283
- justifying reason 15, 23, 24, 25, 32-34, 37, 78, 79, 82, 88-91, 94-96, 98, 106, 109-112, 114, 116-118, 120-122, 125, 126, 132-135, 139, 141, 145, 147, 149, 151, 153, 156-158, 161, 163, 164, 175, 176, 178, 179, 181, 182, 184, 186, 189, 200-202, 204-212, 216, 218, 219, 224-230, 232-237, 239, 242-244, 273, 275, 279, 280, 281
 - objective 33, 34
 - subjective 33, 34, 37, 79, 88, 90, 91, 94-96, 109-112, 149, 152, 230, 242
 - subjective (1) 34
 - subjective (2) 34
- Korsgaard, Christine 9, 26, 27, 37, 42, 76, 81, 98, 100, 110, 130, 133
- McDowell, John 81, 94, 97, 100, 107, 144, 177-181, 185, 198, 213, 214, 250-254, 272, 292, 294
- MF-R Internalism *See* Morality/Reason Internalism
- MJ-R Internalism *See* Morality/Reason Internalism
- Modest Internalism 126, 150, 152-155, 158, 160, 172, 173, 178-181, 184-186, 188, 190, 194, 196-200, 202, 203, 205-207, 210, 211, 214, 216, 219, 225-227, 232, 233, 238, 245
- Moore, G. E. 3, 14, 27, 37, 40-48, 51-55, 60-62, 67, 68, 74-77, 277, 284-290, 293, 294, 298, 299
- moral agent iii, viii, 4, 5, 7, 12, 14, 26, 38, 70, 71, 78, 90, 92-94, 95, 111-114, 116-120, 122-124, 170, 217, 221, 224, 226, 230, 234, 239, 240, 245, 251, 253, 260, 273, 275, 280-283
- moral fact iii, viii, 1, 3, 5-7, 26, 28-30, 38, 39, 77, 78, 92, 95, 124, 224-227, 229-243, 245, 246, 252, 280-283
- Moral Impartiality 4, 8, 9, 10, 14, 26, 78, 88, 92, 93, 224
- moral realism 1-3, 5-7, 9, 14, 26, 27, 29, 38, 77, 95, 225, 245, 246, 283
 - minimal 27
- Moral Realism 1, 2, 4-6, 8-13, 24, 26-29, 37-39, 42, 43, 47, 54, 68, 75, 78, 79, 81, 82, 83, 85, 96, 100, 102, 108, 123, 134, 139, 145, 146, 148, 157, 199, 205, 206, 214, 221, 224, 230, 231, 234, 245, 248, 255, 274, 282, 283, 285-287, 289, 295-300
- Natural 5, 40, 43
- Non-Natural 5
- robust 27
- Supernatural 5
- moral reason 24, 111, 122, 191, 226, 235, 237, 239-241
- Morality/ Reason Internalism 78, 224-226, 229, 231
- MF-R Internalism 229-235, 237, 239, 245

- MJ-R Internalism 229, 230, 232
 Morally Right 243, 244, 246, 273, 275, 279-281
 Morally Wrong 243, 244, 246
 motivation 29, 32, 36, 37, 54, 72, 76, 79, 80-83, 90, 91, 93-100, 102, 103, 109, 115, 125, 126, 140, 144, 153, 158, 160, 165, 190, 203, 209, 215, 222, 223, 225, 296
 Nagel, Thomas 14, 21, 23, 63, 76, 81, 82, 83, 94, 98, 101, 103-107, 293
 natural 1, 5, 27-29, 36, 40, 42, 46, 47, 51, 52, 68, 69, 71, 75, 77, 83, 102, 103, 127, 129, 131, 141, 153, 154, 163, 203, 204, 209, 258, 259, 270, 272, 278
 necessary relation 33, 125, 146, 160, 185, 229, 234
 non-moral reason 114, 121, 233, 237
 non-natural 27, 29, 68
 non-rational agent 111, 113, 114
 non-rational intervention 177-182, 185, 186, 198, 204-206, 218, 219, 226, 282
 objective 9, 23, 26, 33, 37, 38, 67, 77, 230, 248, 251, 254-260, 272, 277
 Open-Question Argument 37, 41, 43, 46, 284
 OQA 42-49, 51-55, 57-61, 63-69, 71-77, 277
 OQA *See* Open-Question Argument
 Parfit, Derek 14, 33, 59, 81, 109, 136, 143, 153, 198-201, 231
 Pettit, Philip 2, 9, 14, 18-21, 23, 33, 64, 84, 100, 127, 134, 138, 139, 145, 149, 200, 244, 246, 248, 257, 290, 294
 possible 4, 11-13, 16, 17, 24, 25, 28, 29, 59, 62, 72, 75, 80, 89, 91, 95, 97, 103, 105, 106, 108, 109, 113, 115, 117, 132-136, 147, 152, 156, 157, 164-167, 170, 171, 173, 175, 183, 187, 188, 190, 191, 201, 203, 212, 216, 219, 221, 222, 225, 226, 229, 237, 238, 240, 243, 244, 252, 253, 268, 279, 280, 282
 cognitively 134, 138, 184
 deterministically 183
 metaphysically 134
 physically 134, 175, 280
 rationally 113, 133, 135, 147, 165, 170, 171, 187, 191, 216, 219, 221, 222, 280
 practical reason 31, 33, 104, 138, 223, 224
 pro tanto reason 34, 153, 199, 229, 233
 properties 3, 10, 12, 13, 41, 42, 46, 47, 53, 68, 70-73, 243, 246-249, 251-255, 257-261, 262, 264-267, 269-272, 281
 R&CN *See* Relative & Contingently Neutral Explanations
 Railton, Peter 2, 26-28, 43, 75, 76, 81, 145, 146, 148, 149, 155, 232, 248, 261, 265, 267-271, 287, 295, 300
 rational agent 15, 16, 21-23, 35, 55, 57, 89, 101, 110, 114, 120-122, 133, 134, 145, 150, 154, 161, 167, 182, 183, 202, 204, 208, 212, 223, 229, 234, 235, 242, 282
 rationality 92, 93, 101, 103, 110, 126, 129, 138, 144, 145, 148, 151, 153, 199, 231, 281
 Raz, Joseph 31, 37, 96, 150
 RD Property 247, 249, 250, 254
 realism 1-3, 5-7, 9, 11, 26, 27, 51, 77, 95, 245, 251, 257, 274, 283, 293, 296
 moral *See* moral realism
 reason
 agent-neutral *See* agent-neutral
 agent-relational *See* agent-relational
 agent-relative *See* agent-relative
 conclusive *See* conclusive reason
 explanatory *See* explanatory reason
 justifying *See* justifying reason
 moral *See* moral reason
 non-moral *See* non-moral reason
 practical *See* practical reason
 pro tanto *See* pro tanto reason
 sufficient *See* sufficient reason
 theoretical *See* theoretical reason
 Reason/ Motivation Externalism 125, 196, 198, 210, 212, 214

- WER 134, 140, 141, 143, 152, 157,
158, 160, 186, 189
- Reason/ Motivation Internalism 77,
79, 125, 126, 133, 139, 146, 148, 149,
151, 155-157, 173, 174, 181, 192, 193,
197, 199, 200, 206, 209, 210, 211, 213,
214, 225
- WIR 125, 126, 133-135, 139, 140,
143-145, 149, 153-157, 180, 184,
188-190, 196-202, 209, 213, 214
- Relative & Contingently Neutral
Explanations 80, 88, 95, 124, 225
- R&CN v, 80, 95-98, 109, 115, 123
- responsible 68, 69, 97, 111, 115, 119,
167, 183, 184, 190, 200, 211, 214, 216,
217-222, 237, 238, 258, 263, 264, 267,
277
- Schroeder, Mark 15-17, 19, 25, 30, 33,
37, 41, 59, 67, 71, 83, 92-94, 97, 107,
116, 118, 125, 139, 149, 150, 175, 190,
199, 208-211, 213, 223, 232, 276, 288,
296, 297
- Shafer-Landau, Russ 9, 28, 43, 54, 81-
83, 85, 92, 93, 96, 100, 102, 103, 108,
133, 134, 139, 157, 174, 199, 203, 205,
206, 208, 210, 211, 214, 231, 245, 274,
287, 288, 291, 292, 296-299
- Skorupski, John 2, 6, 8, 9, 10, 16, 17,
24, 31, 33, 35, 42, 62, 74, 81, 82, 110,
131, 159, 174, 179, 184, 189, 191, 198,
207, 223, 232, 274, 297
- Smith, Michael 26, 27, 31, 33, 42, 54,
60, 61, 64, 68, 69, 72-74, 80, 81, 83-
85, 87, 94, 98, 100, 102, 122, 127, 134,
138, 139, 145-150, 155, 200, 231, 232,
248, 255, 258, 284, 290, 293, 294, 298,
300, 301
- subjectivism 27, 260
- Success Theory 5-8
- sufficient reason 35, 152, 155, 197,
225, 232-235, 237, 239, 282
- supernatural 29, 42, 68
- The Argument from Parsimony 188,
190
- the conditional fallacy 126, 144, 146,
147, 150, 151, 155, 198, 209, 214
- The Explanatory Argument 156, 158,
198, 200, 222
- The Humean Theory of Motivation
38, 79-81, 83, 84, 102, 128, 198, 213,
240, 297, 298
- HTM 77, 79-82, 88, 92-94, 97, 100,
104, 107-109, 115, 213
- The Knowledge Argument 43, 49, 51,
52, 58, 59, 63, 64, 73, 74, 202, 207,
237
- The Reason/ Motivation Argument
79, 92
- The Reason/ Motivation Argument^{II}
224, 225
- The Relative Reasons Argument 38,
78, 79, 88, 94, 95, 109, 123-125, 155,
197
- The Relative Reasons Argument^{II}
224, 227, 230, 232
- theoretical reason 31, 101, 223
- Velleman, J. David 85-87, 92, 117, 118,
149, 198, 199, 299
- WER *See* Reason/ Motivation
Externalism
- Wiggins, David 76, 244, 250-252, 254,
259, 260, 271, 272, 274-277, 300
- Williams, Bernard 10, 22, 33, 77, 103,
125, 127-135, 137-145, 149, 151, 153-
161, 165, 173, 177-180, 186-190, 196,
198, 199, 201-205, 208, 211, 213, 215,
219, 223, 236, 276, 286, 293, 297, 300
- WIR *See* Reason/ Motivation
Internalism
- Wright, Crispin 2, 5, 6, 7, 9, 248, 249,
256, 259, 261, 274, 289, 300, 301

