MORAL INTERNALISM
An Essay in Moral Psychology
MORAL INTERNALISM
An Essay in Moral Psychology

GUNNAR BJÖRNSSON
ABSTRACT

An ancient but central divide in moral philosophy concerns the nature of opinions about what is morally wrong or what our moral duties are. Some philosophers argue that moral motivation is internal to moral opinions: that moral opinions consist of motivational states such as desires or emotions. This has often been seen as a threat to the possibility of rational argument and justification in morals. Other philosophers argue that moral motivation is external to moral opinion: moral opinions should be seen as beliefs about moral reality, beliefs which may or may not motivate depending on whether the person holding them cares about moral matters.

In this essay it is argued that although the traditional case for the internalist position fails, the total available evidence and methodological considerations support an internalist theory formulated in terms of a relatively rich psychological model. It is shown how such a theory can explain not only the practical character of moral opinions and their connection to moral emotions but also phenomena that have been taken to suggest an externalist picture, such as the role of inference, inconsistency, argument and explanations in moral discussion, as well as cases of amoralism and psychological disturbance. In the end, it is concluded that externalist explanations of the same phenomena are methodologically inferior for postulating a more complicated psychology.

Key words: Internalism, moral psychology, moral emotions, objectivism, amoralism

© Gunnar Björnsson
ISBN 91-7153-804-6
Printed by Norstedts Tryckeri, Stockholm
for my parents

and Dr. Hatsumi Masaaki
Acknowledgements

Many are the ill-conceived ideas that would have been included in this essay had not friends, colleagues and teachers come to my assistance with abundant criticism. In particular, numerous errors of innumerable drafts have been exposed during the Thursday seminars at the Department of Philosophy at the University of Stockholm: for that I am grateful to the acute sense of philosophical criticism shown by my supervisor, professor Lars Bergström, as well as to his demands for clarity and his patience with my unfocused production; I am also grateful to fellow graduate students Ulrik Kihlbom, Tomas Månsson, Håkan Salwén and Anders Ullholm for their participation and helpful criticism.

Christian Munthe and professor Torbjörn Tännsjö of the University of Gothenburg have provided helpful comments on parts of the manuscript. Hans Mathlein, Folke Tersman and professor Ruth Millikan, who visited the Department of Philosophy at the University of Stockholm during the spring 1998, all read significant parts of the manuscript and spurred me on to a much needed improvement of the disposition. Simon Blackburn and Michael Smith provided support during a brief visit at Chapel Hill during spring 1997. I thank Simon for his warm hospitality and hours of inspiring philosophical discussion, and Michael for reading a predecessor of section 3.5 and saving me from an embarrassing interpretation of his argument for internalism. I dare not think what this essay would have been like without the support and encouragement of all these people.
Contents

1. Introduction 3
   1.1. A very brief historical survey......................................... 3
   1.2. The amoralist: a call for arguments.............................. 7
   1.3. Preview......................................................................... 8

2. Internalism and externalism 11
   2.1. What kind of theory is internalism?.............................11
   2.2. Moral opinions.............................................................13
   2.3. Motivation.................................................................19
       Strategy-selecting states 19
       Beliefs and desires 24
       Motivation and motives 24
   2.4. Moral optations and practical deliberation.................. 27
   2.5. Moral opinions and moral emotions.............................31
   2.6. Internalism and the language of morals...................... 36
   2.7. Our Choice Internalism............................................... 37
   2.8. The externalist alternative.......................................... 39
       Objectivism 39
       Externalism 41

3. Traditional arguments 43
   3.1. An argument from relativism.......................................43
       Establishing o-relativity 46
       Explaining o-relativity 46
   3.2. More signs of the practical nature of morality.............. 47
       Internalist explanations 49
       Externalist explanations 50
       Concluding remarks 51
   3.3. Why does “ought” imply “can”?.................................... 51
       Two incomplete explanations 52
       A piece of armchair etymology 53
   3.4. Hare’s argument from practicality................................55
       Interpreting the argument 55
       Problems with the argument 58
1. Introduction

Prince: I never thought to hear you speak again.
King: Thy wish was father, Harry, to that thought.
(William Shakespeare, Henry IV, Part 2, IV.v. 92–3.)

In morals, more than in most places, our wishes are fathers to our thoughts: indeed, moral opinions are nothing but wishes or desires. The newly crowned Henry’s opinion that it would be morally wrong of him not to forgive the Lord Chief Justice is the same psychological state as a certain desire to forgive the Lord Chief Justice. And since that is the case, there is no way that Henry could have that opinion without having the desire and hence being at least somewhat motivated to act accordingly. Nothing external to the moral opinion is needed for moral motivation.

“Internalism”, as views positing a tight connection between moral opinions and motivation are commonly labelled, has been both defended and attacked in a number of influential texts within meta-ethics. In this essay I will argue that once we have steered clear of the reefs of uncharitable interpretations and hasty arguments we will see that internalism, although often stated in much too simple terms by both friend and foe, has evidence on its side. Beyond the agitated seas of present controversy lie more fruitful lands of inquiry – or so I hope.

1.1. A very brief historical survey

Moral internalism is by no means fresh from the brains of contemporary philosophy. Two centuries ago, the British moralist Richard Price defended the idea in his Review of Principal Questions in Morals:

All men continually feel, that the perception of right and wrong excites to action; and it is so much their natural and unavoidable sense that this is true, that there are few or none, who, upon having it at first proposed to them, would not wonder at its being questioned.
But further, it seems extremely evident, that excitement belongs to the very ideas of moral right and wrong, and is essentially inseparable from the apprehension of them. [...] When we are conscious that an action is fit to be done, or that it ought to be done, it is not conceivable that we can remain uninfluenced, or want a motive to action.¹

During the same period, Immanuel Kant defended and elaborated the related idea that moral judgements were commands with the logical form of imperatives. But though the idea is old, it apparently retains its appeal and has recently been defended in a number of influential works, such as Simon Blackburn’s *Spreading the Word,*² Allan Gibbard’s *Wise Choices, Apt Feelings*³ and Michael Smith’s *The Moral Problem.*⁴ According to Blackburn, “It seems to be a conceptual truth that to regard something as good is to feel a pull towards promoting or choosing it, or towards wanting other people to feel the pull towards promoting or choosing it”,⁵ and Smith claims that “… having [moral] opinions is a matter of finding ourselves with a corresponding motivation to act.”⁶ On the other side, David O. Brink’s *Moral Realism and the Foundations of Ethics*⁷ contains a much discussed attack on internalism. In between Price and the contemporary debate, defences of internalism and related

⁴ Oxford: Blackwell 1994. Smith provides an argument for internalism but his own theory about moral opinions can be seen as making motivation external to moral opinions. See section 3.5.
⁵ *Spreading the Word*, p. 188.
theories by authors like Charles L. Stevenson, Richard M. Hare and W. D. Falk have been at the focus of the debate.\(^1\)

But long before these writers, and before Price and Kant, David Hume propounded the same idea in his *Treatise*. Not only did he defend the idea, he also drew attention to a possible and troubling consequence of the idea which has occupied philosophers from then on:

> Since morals, therefore, have an influence on the actions and affections, it follows, that they cannot be deriv’d from reason; and that because reason alone, as we have already prov’d, can never have any such influence. Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reasons.\(^2\)

Hume’s sceptical argument has three crucial premisses. The first seems to be internalism: moral opinions are intrinsically motivating. The second is Hume’s famous claim that reason is

---

1 Stevenson: “A person who recognizes X to be ‘good’ must *ipso facto* acquire a stronger tendency to act in its favour than he otherwise would have had.” (“The Emotive Meaning of Ethical Terms” in *Mind* Vol. 46, 1937, p. 16.)

Hare: “I propose to say that the test, whether someone is using the judgement ‘I ought to do X’ as a value-judgement or not is, ‘Does he or does he not recognize that if he assents to the judgement, he must also assent to the command ‘Let me do X’?’” According to Hare, you cannot sincerely assent to a first-person command concerning an action to be done at that time without doing what the command says—which obviously takes some motivation. (*The Language of Morals*, Oxford: Oxford U. P. 1952, pp. 20, 168–9.)

Falk: “...the very thought that we morally ought to do some act is sufficient without reference to any ulterior motive to provide us with a reason for doing it [...] morally good conduct is activated by no other motive than that provided by this thought.” (“Ought and motivation” in eds. Wilfrid Sellars and John Hospers, *Readings in Ethical Theory*, New York: Appleton-Century-Crofts 1952, p. 494.)


“the slave of the passions”, \(^1\) utterly incapable of motivating without the presence of passion of some kind. From these two premisses, it seems to follow that moral opinions are, in part at least, matters of passion or preference. Add Hume’s third premiss that passions cannot be rationally criticised unless they are the results of false beliefs:

A passion ... contains not any representative quality, which renders it a copy of any other existence or modification. ... Tis impossible, therefore, that this passion can be opposed by, or be contradictory to truth and reason; since this contradiction consists in the disagreement of ideas, considered as copies, with those objects which they represent.\(^2\)

Where a passion is neither founded on false suppositions, nor chuses means insufficient for the end, the understanding can neither justify nor condemn it. Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an Indian or person wholly unknown to me.\(^3\)

The conclusion seems to be that moral opinion is beyond the realm of truth, falsity and rational criticism, at least to a large extent. For Hume, that was no fundamental threat to the authority of morality. Although he believed that reason could not by itself yield agreement on moral issues, and although he was well aware of the variety of moral judgement across cultures, he took human nature to be, all in all, a fairly constant and reliable source of moral distinctions.\(^4\)

\(^4\) See the first section of his Enquiry Concerning the Principles of Morals, as well as the amusing and instructive A Dialogue, both in his Enquiries Concern-
Kant, for one, was not so content. He was appalled by the idea that morality should be founded on something as incidental and empirical as the passions of human nature, and repeatedly attacked the premiss that reason should be unable to excite action.1 Others have directed their forces against the premiss that passions cannot, in themselves, be irrational. In this essay, however, we shall primarily be concerned with the tenability of the first premiss of Hume's sceptical argument; the tenability of internalism.

1.2. The amoralist: a call for arguments
Most of us expect moral opinion to be accompanied by relevant motivation, but critics of internalism — “externalists” — claim that there are, or can be, moral opinions unaccompanied by motivation. Those who have such opinions — the “amoralists” — are the principal characters in the most frequently invoked argument against internalisms of various forms.

The amoralist argument starts with either the weak claim that we can imagine the existence of amoralists, or that it is conceivable, or with the stronger claim that there are amoralists; and it further claims that this could not be if internalism was correct.2 But without further discussion, such arguments do

---

1 See for example Grundlegung zur Metaphysik der Sitten, books 2 and 3, and Critique of Practical Reason. Kant famously claimed that to be rational we have to act under the idea of freedom, taking ourselves to be in a position where we can will not to act from our desires. For Kant, the only way to do that is to act according to a maxim of which we can at the same time will that it becomes a universal law. (Grundlegung zur Metaphysik der Sitten [1785], p. 52. References are to the second German edition, the pages of which are indicated by marginal numbers in Paton’s translation The Moral Law, third edition London: Hutchinson & Co., 1956.)

2 Michael Stocker argues this way against moral internalism in “Desiring the Bad: an Essay in Moral Psychology” in The Journal of Philosophy Vol. 76, No. 12 1979, see e.g. p. 742. In Moral Realism and the Foundations of Ethics, David O. Brink argues in the same way on pp. 27, 46–9. One of Allan Gibbard’s critics, Richmond Campbell, objects to Gibbard’s internalism on these grounds on pp. 316–8 in his critical notice of Gibbard’s Wise Choices, Apt Feelings in the
not take us very far: peoples’ imaginations and their interpretations of imagined or allegedly real amoralists are at best fallible guides to the constitution of moral opinions, and the difficulty is aggravated by the fact that a number of people with a reputation for intelligence and wisdom have declared themselves unable to imagine a true amoralist.¹

For the amoralist argument to do more than remind us that the internalist thesis is controversial, we must be given reasons to believe that the critic’s intuitions and interpretations are more veridical than the internalist’s. Further arguments are needed, and it is these arguments rather than reports of amoralist imaginations that could refute or support internalism.

1.3. Preview

The bulk of this book is an examination of the most important arguments that have been put forth for and against internalism, conjoined with a presentation of what I take to be a decisive argument for internalism at this stage of inquiry. But even though the formulation of internalism given at the beginning of this chapter goes a long way to define the present field of inquiry, more needs to be said to prepare the ground for that examination. My hope is to avoid the often cumbersome and sometimes tedious situation where a considerable number of related theses are discussed at the same time, making it difficult

¹ See for example Jonathan Dancy’s reply to the amoralist challenge in Moral Reasons, Oxford: Blackwell 1993, pp. 4–6. Allan Gibbard’s “Reply to Sinnott-Armstrong” in Philosophical Studies Vol. 69, 1993, is in much the same vein (see pp. 318–9).

For an enlightening work on the methodology of thought-experimental arguments, see Sören Häggqvist’s Thought Experiments in Philosophy, Stockholm: Almqvist & Wiksell 1996.
for the reader to keep track and nearly impossible for her to stay interested as she makes her way through variations of doubtful significance.

In the next chapter, then, we shall select the ingredients that will make up our choice internalism. To put it briefly and in terms that will need explication, internalists should claim that a moral opinion to the effect that some act is wrong is an optation\(^1\) (desire or wish) that the act is not performed: this kind of optation results from a species of practical judgement and its workings are intimately connected to emotions such as anger and guilt. Externalism, by contrast, denies that moral opinions consist, even in part, of optations: rather, moral opinions about an action are seen as beliefs representing moral facts about that action, facts that are independent of the optations of the person having the opinion.

With the objects of our investigation defined, we will be prepared for a journey through the argumentative landscape that has taken form in meta-ethical debate. In chapter 3, it is argued that traditional support for internalism fails. In defence of their theory, internalists have invoked the observations that we expect moral opinions to be accompanied by motivation in various contexts; that we form moral opinions almost exclusively about what we take to be performable actions; and that we tend to utter moral sentences in contexts where the point of making the utterance is to affect action. Here it is shown that the externalist is perfectly able to explain these observations, given a number of auxiliary hypotheses that seem reasonable under the externalist assumption that moral opinions represent objective moral facts.

That objectivist assumption is examined in chapter 4. A number of observations that have been put forth in support of

---

\(^1\) According to the OED, an optation is an “action of wishing; a wish or desire” (or an expression thereof) or a “choice or preference”. Here, it shall just mean “wish or desire”. (I thank Per–Erik Malmnäs for bringing “optation” to my attention.)
objectivism are discussed but it is concluded that all of them can be explained from an internalist perspective. Given the assumption that moral opinions are moral optations, we can understand – as well as we can under externalism – why it is that moral opinions share many traits with paradigmatically objective beliefs. For example, it can be explained that we make inferences involving various kinds of moral opinions and that we explain events employing a moral vocabulary, but also that we feel that moral opinions can be mistaken, that moral disagreement is something more than mere differences in desire, and that many acts are wrong independently of what we happen to think about their moral status.

Since chapter 3 has revealed reasons to assume that moral optations are an important part of psychological reality (whether or not they are part of what constitutes moral opinions), and since our discussion of arguments for objectivism will have given us no reason to assume that they are anything more than moral optations, we will have the beginnings of an argument for internalism.

But it might be thought that there are obvious cases where moral opinions come without the right kind of optations, cases of amoralism and psychological disturbance. Accommodating such intuitions in an internalist framework is the business of chapter 5. Here it is shown how internalism can explain why we have such intuitions, but it is also argued that familiar cases where moral opinion and optation are believed to come apart really do involve the relevant moral optations. This argument, together with the results of chapters 3 and 4, vindicates internalism. Unfortunately, the way we have to travel to enjoy that view demands long detours around argumentative and methodological pitfalls and careful bridging of explanatory gaps. At this stage I can only hope to have tempted the reader to follow my trail.
2. Internalism and externalism

In this chapter, we will consider the various ingredients that make up the subject matter of the ensuing discussion: our choice internalism. The strategy is to choose an interpretation that has some prospect of being true while being strong enough to be of significant philosophical interest, should it be true.

To select one fairly specific form of internalism will obviously leave untouched many of the versions of internalism that are found in the literature: at least there will be no explicit discussion of their pros and cons in chapters 3 through 5 where we make a thorough survey of the arguments found in the literature. My belief, though, is that compared to our favoured internalism, each of these versions is either (a) so similar to it that most of what is said in this essay will pertain to that other version as well, or (b) so dissimilar to it that a full discussion of it would demand another book, or (c) less plausible or philosophically interesting. In order not to encumber the book with comparisons of numerous alternative interpretations, I leave for the reader to decide whether this is true.

2.1. What kind of theory is internalism?

When internalists claim that if a person has the opinion that it would be wrong of her not to perform a certain action, then she will be motivated to perform that action, the implications could be of various kinds. What many would agree to – even self-proclaimed opponents of internalism – is that if a person has the opinion that it would be wrong of her not to perform a certain action, then probably she will be somewhat motivated to perform that action. An interesting form of internalism makes a stronger claim, but which? Some authors take internalism to be a conceptual truth, a truth about the meaning of expressions such as “Lisa is of the opinion that it would be morally wrong to administer capital punishment”; for some this truth is supposedly discovered in an analysis of our use of such expressions, for others it is the result of stipulation. Others seem to think that
internalism is a theory about metaphysical necessity, whatever that might be, while others still seem to conceive of it as an empirical hypothesis about natural necessity, a hypothesis to be supported by psychological and perhaps biological theory.

The version of internalism to be discussed in this essay will be related to conceptual or linguistic issues as well as speculative psychology. Put a bit roughly, it claims that:

1. There is a kind of psychological state — “moral motivation” — details of which are added in this chapter.

2. Peoples’ opinions about the moral rightness or wrongness of actions — their “moral opinions” — consist of moral motivation.

Conceived of in this way, internalism is on a par with claims familiar from empirical science, such as the claim that water is constituted by a compound of oxygen and hydrogen, or that temperature is mean molecular kinetic energy. In other words, it is not a claim to be settled a priori. For example, claim (1) invites criticism and support from psychological theory and observation, and the characterization of moral motivation given in this chapter aspires to a certain psychological realism. Moreover, (2) makes relevant evidence that has been invoked for and against analyses of moral language, since it is hard to see what our opinions about moral rightness and wrongness could be if not the kind of psychological states we typically express when we make moral statements, that is, when we use moral language. (See sections 2.2 and 2.6.) Nevertheless, (2) does not explicitly say anything about moral language: it is a claim about certain psychological states — moral opinions and moral motivation — and as such open to refutation or vindication by observation of and theorizing about these states, not just with reference to the workings of moral language.

The bet is that this interpretation of internalism makes a variety of arguments in the literature relevant without adding anything obviously implausible to what most internalists have
said, and without subtracting any factor contributing to the appeal (or horror) of Hume’s anti-rationalist argument.

2.2. Moral opinions

When internalism is discussed in the literature, the kind of entity that it is taken to concern is sometimes referred to as a “moral judgement” or “ethical judgement”; as a “value judgement” or “normative judgement”; as an “ought’-judgement”, “judgement of obligation”, or “judgement of duty”, or as a “moral thought”, “moral consideration” or “moral opinion”. Sometimes, yet, judgements to the effect that one course of action is better than another are the subject of the debate. Often writers put the issue in several of these ways, without paying much attention to the different directions in which they might lead us.¹

But not paying attention to the cross-roads leaves us lost, not knowing why others ended up at other places, and frequently believing that they went off the road into a jungle of misguided thoughts. Not every kind of judgement is constituted by moral motivation and it is entirely conceivable that we could end up with internalism for “moral judgements” or “moral opinions” but not for “normative judgements” without having made any mistake, and without being involved in contradictions. All we have to suppose is that the various labels in use apply – as they seem – to somewhat different kinds of entities.

To reduce such difficulties, we shall say that moral internalism (or “internalism”, for short) is the view that:

(A) Opinions to the effect that it would be morally wrong of one to perform a certain act are constituted by moral motivation not to perform that act. (This claim presum-

¹ See for example Charles Stevenson’s Ethics and Language. New Haven: Yale U. P. 1944, where “moral judgement” is used on p. 13, but “ethical judgement” in most other places; Wilfrid Sellars “On knowing the better and doing the worse” in the International Philosophical Quarterly Vol. 10, 1970; W. D. Falk “Ought’ and motivation” and Wilfrid Sellars “Obligation and Motivation”, both in eds. Sellars and Hospers, Readings in Ethical Theory.
ably includes opinions to the effect that an act is morally forbidden, or that – morally speaking – it must not be performed, or that it is one’s moral duty not to perform it.)

(B) Opinions to the effect that it would be morally wrong of one not to perform an act are constituted by moral motivation to perform it. (This claim presumably includes opinions to the effect that it is one’s duty to perform an act, or that it is what one – morally speaking – must perform.)

Our label for such opinions will, as a matter of stipulation, be “moral opinions”.¹ Note that many opinions that might intuitively be thought of as moral fall outside the scope of internalism: opinions to the effect that a person or act is good or bad, admirable or abominable, virtuous or vicious, but also (for the time being) opinions to the effect that an act is morally right or morally permissible.

At this point, some readers might wish for a definition of “opinions to the effect that (not) performing some act is morally wrong”. But since it is hard to see how we could supply such a definition without taking a stand on exactly the thesis under examination, we will have to start with something less definite than a set of informative necessary and sufficient conditions. The most sensible approach is to consider paradigmatic or uncontroversial cases where people have the opinion that it would be morally wrong to do this or that, and ask whether these opinions are constituted by motivational states. Still, in order to clarify the issue we shall consider some of the means that we employ when identifying moral opinions.

Most often, paying attention to what people say is the only workable way we have of identifying the moral opinions of others: a person will typically have the opinion that abortion is

¹ Moral opinions are neither judgements in the sense of mental or linguistic acts, nor abstract entities (“propositions”), but psychological states.
morally wrong if she says that abortion is morally wrong, and our foremost way of identifying our own moral opinions might well be to pay attention to what we think, to our “inner” talk, and to our thoughts about what we would be ready to say.

Important as these procedures are, however, the fact that we engage in pretence and acting every now and then, both in speech and thought, makes them fallible. We identify make-believe by context (a play, for example, or a thought-experiment) or by intonation, and we often recognize pretence by signs of the lack of desire to act in accordance with what has been avowed.¹ Such lack, in turn, is sometimes recognized by observing the actions and the emotional expressions of the interpretee.

But the linguistic procedure has problems even when utterances are perfectly sincere and made in the speaker’s or thinker’s own voice. In some contexts we are not tempted to say that a thought or utterance expresses a moral opinion even if cast in terms what is right, wrong, or what must or ought (not) occur: as when I tell myself that I ought to put more salt in the stew, or am told that it was wrong to move the pawn. The question, then, is how we distinguish moral rights, wrongs, musts or oughts from their non-moral counterparts. And since it is quite rare that speakers prefix their statements about what ought or must be done or what is right or wrong with “m Morally”, our means of distinguishing moral opinions from other psychological states expressible by such statements must go beyond the identification of the relevant vocabulary and the overt syntactic form of the utterance.

I doubt that we have very clear and uniform intuitions about where to draw the line between moral and non-moral uses of “ought”, “wrong” and “must”, but it does seem to have something to do with the kind of considerations that are taken

¹ This does not in any straightforward way imply that the psychological state that we identify consists of a desire. (This claim is supported in section 3.2.) It is both implausible and unnecessary for opponents of internalism to deny that lack of relevant motivation suggests lack of moral opinion.
when forming the opinion. If only considerations relevant to economic results or victory in sports are taken, say, the resulting judgement will probably not be thought of as a moral one, even if it is expressed in terms of what it would be wrong to do or what one must do. As a first approximation, then, we could perhaps say that we identify a moral opinion by recognizing that its formation involved considering such things as the effects of the act in question on peoples’ welfare and whether it is an act of respect or disrespect.¹ (Note that the issue discussed at this point is epistemological rather than ontological: it focuses on our means of identifying moral opinions rather than on their nature.) Although it is difficult to sharpen this approximation into a theoretically pleasing demarcation of moral opinions, we do seem to have rough ideas about the kind of reasons that should elicit a state of mind in order for it to be a moral opinion. If an opinion does not have grounds roughly belonging to that kind we are reluctant to say that it is an opinion about the moral status of an act.

Allan Gibbard objects to a similar thesis,¹ arguing from the following two observations:

1. Personal burdens and opportunities do not count as moral considerations when weighed against other non-moral considerations, but they seem able to defeat moral considerations, making it morally right to do something that would otherwise have been morally wrong.

2. We sometimes ask whether or not to be moral, and on such occasions we seem to be weighing personal burdens and opportunities against moral considerations.

The problem, according to Gibbard, is that on the first observation the result of weighing together personal burdens and moral considerations is a moral judgement, whereas on the second observation it might be a judgement to the effect that one should not do what is morally right. But then it must be something other than the considerations involved that decides whether (we would say that) the resulting judgement is moral, Gibbard suggests: the same considerations are involved in both cases.²

As far as I can see, this is not a very difficult problem for the view sketched above, although it calls for some clarification. Gibbard himself seems to recognize that it is when personal enjoyment defeats considerations which would otherwise result in a slight moral obligation that the conclusion might be that, all told, it is morally right to go for enjoyment or avoid burdens.³ It is when we ask whether we should give more weight than that to our own projects that our question is whether to be


² Gibbard’s alternative theory – which will be considered in section 2.5 – is that moral judgements are judgements about whether the act in question should invoke moral feelings.

moral. This suggests that considerations are identified as moral in virtue of the way in which they affect judgement and not only in virtue of their representational content. Of course, this is quite obvious once the issue is raised – perhaps too obvious to be noticed. For example, if someone lets the consideration that an act would be a breach of promise affect his judgement about that act, but as a reason for thinking that the act is obligatory, a resulting opinion that the act is obligatory (“this is what I must do”) or that it is wrong (“this I can’t do even though it is a splendid breach of promise”) would presumably not be thought of as an opinion to the effect that it is morally obligatory or morally wrong. Similarly, it is not the mere fact that someone weighs personal enjoyment against truth-telling, say, that determines whether we think of the resulting opinion as moral: the relative weights he gives must be in the right neighbourhood.

A related feature illustrated by Gibbard’s examples is that a consideration is regarded as moral or non-moral only in the context of an overall assessment of the act that elicits a moral opinion. That a consideration of personal benefit fails to qualify as a moral consideration by itself does not mean that it cannot be a moral consideration when it is part of a suitable overall assessment, for example when lying in some trivial context is weighed against the personal costs of honesty.

With these qualifications, it is reasonable to assume that we often identify moral opinions by identifying the set of considerations that are given weight in the deliberation and what weight they are given. To be sure, just about anything can be a sign of a moral opinion if the context is right. But linguistic expressions and the kind of considerations that we make are probably our most important means of identification, together with the expression of “moral” feelings such as guilt or indignation. (More of them in section 2.5.)

What internalism claims, then, is that there is a kind of psychological state that we identify by the means sketched above, and that this kind of state consists of moral motivation.
This claim is similar to the claim that the kind of liquid that we call “water” and identify by sight, smell, taste, and so forth consists (almost entirely) of HOH-molecules. Of course, the concrete methods suitable for ascertaining the chemical constituents of a liquid are not the same as those apt for analysing a psychological state. In the next section we look closer at the sorts of constituents that we are looking for, and how we can, or cannot, expect to identify them.

2.3. Motivation

When a person is of the opinion that it would be morally wrong of her not to take a certain route of action, that opinion is constituted by motivation to take that route — or so internalism has it. But what is it to be so motivated? And how can we decide whether the relevant motivation is present? Unless we are quite clear about this, it will be difficult to draw the relevant implications from internalism, and thus difficult to put the theory to test. In this section, therefore, we will try to bring some light to these issues.

Some readers of the following paragraphs might get the impression that the model or theory of motivation presented here is redundant because it goes little beyond folk-psychology. But since the terminology of folk-psychology is less than clear in many cases and since it seems that much of the controversy over internalism is an effect of having only implicit or rough models of motivation, it is necessary to explicate the model and the assumptions that will be employed throughout the essay. If nothing else, that should make it easier to level accurate criticism at the argument to be pursued in the following chapters.

Strategy-selecting states  As far as I can tell, the most plausible theory about the relation between motivation and action is cast in functional terms, saying that a person is motivated to perform a certain act when she is in a state the function of which is to elicit that performance. Of course, a full and proper assessment of this approach demands explanation of what it means
that something is the “function” of a thing: a variety of analyses of that concept can be found in the literature, and as Larry Wright reminds us, the term “function” serves several uses in various fields.¹ The bet here, however, is that some reasonable sense of “function” would put the following story close to the truth.

In the guise of mutation, natural selection and inheritance, Mother Nature has bestowed on us systems that have the function of selecting responses to the environment, by (1) representing or mapping² features of our environment, our bodily states and the kind of things that can be done with the body in the environment, and (2) selecting behavioural strategies in response to these representations.³ (Note that since maps can be

² Our inner maps need not look anything like your standard tourist map of the city centre, of course: indeed, physical resemblance between map and mapped need not play any role in the story. The important thing about maps is that they can serve their proper function in a purposeful way only if features of the map correspond to features of the environment. For a more careful and detailed account of mapping from a biological perspective, see Part II of Ruth Garrett Millikan’s *Language, Thought, and Other Biological Categories*, Cambridge, Mass.: The MIT Press / A Bradford Book 1984.
³ In a way, the ultimate biological function of any such system is to promote the propagation of its bearer, or more correctly, of its genes. (See Richard Dawkins’ *The Selfish Gene*, Oxford: Oxford U. P. 1976.) That statement of function is true, and interesting, but more like the most important background assumption in any investigation into biological function than its result. (See Daniel Dennett’s *Darwin’s Dangerous Idea*, New York: Simon & Schuster 1995 for a recent defence of that view.) When we want to know the biological function of some trait we want to know in what way that particular trait contributed to biological success of its bearer or its genes, not that it contributed, or in what way all traits contribute. See Ruth Garrett Millikan’s *Language, Thought, and Other Biological Categories*, pp. 35–6, and Karen Neander’s “Misrepresenting and Malfunctioning” in *Philosophical Studies* Vol. 79 1995, pp. 109–141, for a discussion of how relevant function should be understood.

I agree with Dennett (The Intentional Stance, Cambridge, Mass.: The MIT Press / A Bradford Book 1987, and *Darwin’s Dangerous Idea*) that ascription of function is a fairly loose business, following particular explanatory needs. For discussions of indeterminacy in the ascription of motivation, see also Donald Davidson’s *Essays on Actions and Events*, Oxford: Oxford U. P. 1980, e. g. essays 11 and 13 and *Inquiries into Truth and Interpretation*, Oxford: Oxford U. P. 1984, e. g. essays 10 and 16.
incorrect and strategies fail to be implemented, the relevant notion of function must allow for *malfunction* as well.) The former systems are what allow us to have beliefs about the world, the latter are what allow us to act on such beliefs.\(^1\) For example: when I bought my friend a drink the other day, I did so because:

(1) I believed that it would make her feel better — without that belief, I would not have bought that drink. That is, I bought my friend the drink because I represented the situation as being one in which buying her a drink would make her feel better.

(2) I desired to make her feel better — if I had been indifferent to her feelings I would not have bought her that drink. That is, I bought her that drink because my response selecting systems were in a state the function of which was to make me do X on condition that (I believed that) X would make her feel better.

The state incarnating the strategy to make my friend feel better was *potentially* activating, triggered by the representation of an action that would make her feel better. Such states, the function of which is to elicit action when triggered by the relevant representations, will be referred to as “strategy-selecting states”.

Sometimes, of course, two or more conflicting strategy-selecting states are triggered at once. Perhaps, in the above example, the representation of the meagre contents of my wallet and the

---

\(^1\) G. E. M. Anscombe described the case where a man intends to buy everything that is on his shopping list, and is being followed by a detective intending to write down all that the man buys. What the man’s list is supposed to accomplish is to make it the case that everything on it is bought — to make the world fit, whereas the detective’s list is supposed to contain all that the man buys — to fit the world. (*Intention*, Oxford: Blackwell 1957, p. 56.) In some sense corresponding to the relevant sense of “function”, strategy-selecting states are “supposed” to cause the world to be in a certain way, whereas representational states are “supposed” to reflect the way the world is.
price of the drink triggered a state the function of which was to elicit responses not to make me broke, causing motivation not to buy that drink. On this occasion, though, precedence was taken by the motivation elicited by my belief that buying my friend a drink would make her feel better.

The activating function of strategy-selecting states, then, is typically conditional in at least two ways. First, the state must be triggered by some representation of the action and the situation; second, it must not be in conflict with some “stronger” triggered strategy-selecting state. We might say that the strengths of strategy-selecting states reflect the flexible and frequently adjusted strategic priorities of the response selecting systems. In other words, the response selecting system creates strategy-selecting states the function of which is to elicit action on condition that the action is represented in the relevant way and no stronger, conflicting strategy-selecting state is triggered at the same time.

Admittedly, nothing much has been said about the details and particular workings of our strategy-selecting and representational systems. But leaving the details of the matter, the functionalistic internalist can hold that although being of a certain moral opinion is to be in a certain strategy-selecting state the function of which is to elicit (or prevent) action, this is not, in turn, to act in any special way on that occasion.1 So, although

---

1 Any detailed theory of the workings of our representational and strategy-selecting systems will have to invoke the results of speculation about and inquiry into the machinery of our nervous and hormonal systems. The epistemic situation can perhaps be illuminated if we consider the portable computer on which I am writing. In response to the input that it receives from the keyboard and in response to the information sent between its various routines it can perform a number of tasks, including displaying the text that I am feeding it, making automatic back-ups of my files, updating an On Location™ catalogue of every word stored on the hard disc, driving a portable printer, and so forth. But it cannot perform all these tasks at once, especially not when batteries are low, and so it has to give the various tasks different strategic priorities, putting some on hold – perhaps indefinitely – while performing others, and there must be states in the computer that are causally responsible for initializing or withholding the various tasks. Personally though, I have no idea how this
the way a person acts will tell us much about her motivation, it does not reveal it in any simple and straightforward way. Still, our phenomenological access to strategy-selecting states exceeds that of observations of performed action. In many cases when strategy-selecting states have been triggered but do not, for the moment, elicit action, we can still feel them working: how they feel will depend on the violence and saliency of changes in the neural and bodily states that they involve, cause, or are caused by, as revealed by the various degrees and kinds of introspective access that we are familiar with.¹

It is not hard to imagine why strategy-selecting states should be somewhat open to introspection, since it gives us opportunity to modify or suppress them or adjust our behaviour in order to be able to act on them more efficiently. Realising that you desire the company of some person can give you an opportunity to consciously arrange your life to be able to be with him or her, or to be unable to be with him or her if that company would be a bad thing, or perhaps decide not to let that desire have any effects on your actions. When we say that we want to do something that we nevertheless refrain from doing because we want something else more strongly or because the opportunity never comes up, our ground for saying so is often this kind of introspective access.

¹ In his book *Descartes’ Error* (G. P. Putnam’s Sons 1994), Antonio Damasio provides interesting examples and theories about (among other things) strategy-selecting states, and how they are affected by neural damage. See especially Part II of that book but also section 2.5 of the present study for the connection between strategy-selecting states and emotions.
Beliefs and desires  Above, we have used “belief” and “desire” as more or less synonymous with “representational state” and “strategy-selecting state”. But it should be stressed that not all representational states are beliefs, and not all strategy-selecting states are desires. The representation of bodily temperature involved in the regulation of sweating would not be classified as a belief that the body temperature is too high, and states the function of which is to cause sweating in response to those representations are not desires to sweat. We might perhaps say that the body believes that it is too warm, and desires to lower the temperature by sweating, but ascribing beliefs to a person seems to imply that the states in question are comparatively available for theoretical and practical reasoning of a kind relevant for what we think of as voluntary action rather than mere reaction. Similarly, desires seem to be strategy-selecting states that elicit voluntary action rather than any kind of behaviour; at least, desires so conceived are the subclass of strategy-selecting states that we need for our characterization of internalism. This definition of beliefs and desires is very rudimentary of course, and at least as vague as the concept of voluntary action, but I believe that it leads our thoughts in the right directions.

In the internalist debate, it is sometimes said that to have a moral opinion is to have a desire to act accordingly, and that giving voice to a moral opinion is to express such a desire. Without qualification, however, the word “desire” sometimes suggests a strength and a phenomenological saliency that might render internalism implausible – which is the reason that our original statement of internalism was put in terms of motivation rather than desires. It should be noted, therefore, that when “desire” is used in this essay, it refers to the kind of strategy-selecting state especially relevant to the causation of voluntary action, without any special implications concerning phenomenology or strength.

Motivation and motives  The view of desires presented above belongs to a tradition finding its first ardent advocate in Hume.
Hume’s “passion” is the forerunner of our “desire” and his “idea” the forerunner of our “belief”.1 Jonathan Dancy attacks this tradition, objecting to the view that every action is the product of beliefs and desires, and such an objection would seem to be at odds with what has been said here.2 But when Dancy is to explain his complaint, he writes that what he is “in the business of denying is that the desire must be (a part of) what motivates the agent.”3 We should grant Dancy that we seldom refer to desires as that which motivates us. (My motive or reason for buying my friend a drink was that she needed one, not that I wanted to give her what she needed.) But that seems to be a terminological point, and of little importance once it is appreciated as such.4 It does not contradict the fact that the presence of a desire is a necessary prerequisite for motivation and voluntary action, even when we are reluctant to say that the desire is part of what motivates us. In order for someone to perform a voluntary action she must be motivated to perform it, and in order for her to be motivated to do that it is insufficient to have a representation of the act, or even to be caused to act by such a representation: unless something (a desire) makes it the function of the representation to cause action there is no


3 Ibid.

4 This is also the view of Robert Noggle, who argues that the debate between cognitivists (such as Dancy) and conativists (such as Smith) is based on little more than a disagreement in vocabulary. See “The Nature of Motivation (And Why it Matters Less to Ethics than One Might Think)” in Philosophical Studies Vol. 87, No. 1, July 1997.
reason to say that it (or the fact it represents) *motivates* the person in the sense relevant for voluntary action.¹

To avoid terminological quibbles and to connect what has been said with the internalist issue we shall turn from “desire” to expressions such as “motive”, “motivation”, and “being motivated” – the latter being the crucial expression in our previous formulations of internalism. When we say that someone is motivated to perform some act, X, we shall mean that a belief about X has triggered a strategy-selecting state (which may or may not be a desire) the function of which is to cause action of the represented kind, thus eliciting a desire to perform X. “To have motivation” is taken to mean the same as “to be motivated”, and sometimes we will use “motivation” instead of “desire”. So, when internalism says that

if Henry is of the opinion that it would be morally wrong not to forgive the Lord Chief Justice, he is motivated to forgive the Lord Chief Justice,

it should be understood as saying that

if Henry is of the opinion that it would be morally wrong not to forgive the Lord Chief Justice, a strategy-selecting state of his has been triggered by a belief about the act, eliciting a desire to forgive the Lord Chief Justice.

Before we turn to other issues, it should be acknowledged that saying that someone is motivated to do something is ambiguous. At times it communicates that the person is somewhat attracted by the act, at other times that she is more attracted by the act than by the relevant alternatives. It should be clear from our definitions that it is the former notion that we have in mind here.

2.4. Moral optations and practical deliberation

An internalism put in terms of motivation will be mute on most moral opinions concerning the actions of others, or actions that have already been or cannot anymore be performed: such acts are the proper objects of other conative attitudes (wishes, in particular) but not of motivation. But it might seem that moral opinion concerning the actions of other people also consists of motivation or attitudes. It might be held, for example, that we are always (somewhat) motivated to prevent others from doing what is morally wrong according to our opinions, and that we always wish for actions that we judge to be morally wrong but outside our control not to be (or not to have been) performed.

Before we decide what internalism should say, we shall note a few connections between some of these suggestions. First, it is helpful to remember that when we learn to form and give reasons for our moral opinions, we typically learn that if it is morally wrong (or obligatory) for you to do X, then it is equally morally wrong (or obligatory) for me to do X in exactly the same circumstances: the mere fact I am I is not a reason for making moral distinctions between my actions and yours. In other words, our reasons for moral opinions are typically beliefs that are in a certain sense impartial, and the corresponding motivation to act accordingly is in a corresponding way impartial: what motivates, morally, is some fact about the action that could, in principle, be a fact about the actions of anybody, at least of anybody in one’s own society.\(^1\)

It should also be held in mind that wishes and desires / motivation are intimately connected. Indeed, wishes seem to be strategy-selecting states just as desires, although further removed from action: if I wish that something was the case and an opportunity to make that something the case appeared (without great costs tied to it), I would (if functioning properly) be motivated (or desire) to make it the case. Similarly, if I

---

desire that something will be the case but learn that it will not be the case regardless of what I do, I will still wish that it would become the case.

So, given the fact that I am motivated to perform a certain act because I believe that it has a certain impersonal characteristic, it seems reasonable that I should wish for the performance of other acts (outside the domain of my own performable acts) having the same impersonal characteristic. In other words, it seems that in order not to make arbitrary distinctions, internalism should say not only that moral opinions about one’s own case consist of motivation, but also that moral opinions about the actions of others consist of corresponding wishes. Generally put: internalists should claim that moral opinions – whether concerning actions within or outside one’s power – are moral optations (desires or wishes) directly elicited by a belief about the act in question. Moreover, given that I could make it the case that someone else performed an action with the relevant impersonal characteristics, it can reasonably be expected that I should be (somewhat) motivated to make him perform it, although it is likely that other factors will tell in favour of not intervening, and thereby counteract such motivation. (Sometimes we wish that someone would do something – notice us, remembers our birthday, keep a promise – without being made to do it. In such cases, of course, we never find the opportunity to act so as to satisfy our wish.1)

If internalism is correct, then, moral opinions are optations of a kind that we often identify by the considerations that gave rise to them (as discussed in section 2.2). But as we will see, the internalist should not take all optations elicited by such “moral” considerations to be moral opinions. Suppose, for example, that Todd and Chip disagree about whether to punish their nephew, Ned, for stealing apples from the neighbour. Both realize that the boy will suffer if he is punished and both have

1 In section 5.5 we shall see how agent-relative moralities might be partially constituted by such wishes.
the optation that he should not suffer. Furthermore, both realize that punishment will have educational effects, and both have the optation that such effects take place. The grounds that Todd and Chip have for their optations presumably qualify as moral: had a judgement been made on such grounds it could well have qualified as a moral judgement.

So far, there is no disagreement or indeed difference between the two with respect to their moral optations. But whereas Todd wants to give weight to suffering when deciding whether to punish or not, Chip does not, at least not when the suffering is moderate. And whereas Chip wants to give weight to educational effects even when suffering is at stake, Todd does not, at least not when the educational effects are moderate. The result is that Todd forms the considered desire that they do not punish Ned while Chip forms the opposite considered desire. Here, internalism should identify their moral opinions with these latter desires rather than with, for example, Todd's remaining wish that Ned would learn to behave or Chip's remaining wish that Ned would not suffer. In general, moral opinions should be identified with the results of practical deliberation concerned with moral considerations.

Here it is important to take "practical deliberation" in the right way. The central case, of course, is when one tries to decide what to do, but the same psychological mechanism can be applied when we consider the actions of others, or actions that we once performed, or actions that we might at some time be in the position to perform. (The latter kinds of deliberation prepare us for blaming, praising, justifying, giving advice, and for future action. See sections 2.5, 4.4 and 4.7) The suggestion, then, is that practical deliberation is a natural psychological capacity and that moral opinions / optations are its results when it has been concerned with moral considerations.

As with any practical or theoretical judgement, it is possible to remember its results without remembering its grounds, and this turns a person's moral opinions into a potentially messy bunch. Some results of moral deliberation might be more stable
in the sense that the moral judge has a stable tendency to form the same moral opinions when reconsidering the action or kind of action in question; others are less stable because they are formed in extraordinary situations or moods; and any opinion might come into conflict with remembered results formed under different conditions. For various reasons, we might be prone to resolve such conflicts in one way rather than another: we might want, as classical utilitarian hedonists have wanted, to form our moral opinions under consideration of all and only the consequences of an action and its alternatives for sentient beings, or we might want to form them under consideration of the kind of beings that we would be if we acted thus and not so. Or, realizing the utility of allowing moral opinions to be formed differently at different times— in more intuitive fashion when time is short, but more reflectively on our philosophical Sundays, say—we might nevertheless choose to advocate, express and nurture those formed in one style for some other reason: for example

---

1 An important part of our capacity to judge situations consists in a capacity to consider certain aspects of a problem rather than others. In the most simple case, it is a matter of not looking or listening or smelling because one expects that the impression will distort one’s decisions, as when we avoid looking down when climbing a high ladder, avoid breathing through our noses when faced with foul smells, concentrate on what the person in front of us says rather than on the music to our left, or look at one map rather than another. Similarly, we can concentrate our thinking and imagining on certain things rather than others. Sometimes this is more deliberate than at other times, and a shift in focus can be caused by such a variety of things as perception of extreme danger, a slap in the face, the thought of a parent, a loved one, or God, the thought that one will have to justify one’s actions to others, or a desire (not) to act from a certain kind of desire—a “Kantian” desire not to act from desires that one cannot will that others act from, say, or a (romantic) desire to act from desires formed under the influence of strong emotions—in conjunction with the belief that one is (not) about to act from a desire of that kind. I agree with Frederic Schick’s *Understanding Action*, Cambridge: Cambridge U. P. 1991 in taking people’s capacity to see things from different angles or under different conceptions as central in understanding their actions. Much earlier than Schick’s application of this idea to decision theory, Iris Murdoch stressed the moral importance of our capacity to see things in different ways. See her “The Idea of Perfection” in *The Sovereignty of Good*, London: Routledge and Kegan Paul 1970.
because they are less likely to be inconsistent with one’s other
opinions or more likely to be what others would approve of.¹

2.5. Moral opinions and moral emotions
Feelings or emotions do not plausibly constitute moral opinions: moral opinions are, in general, much more persistent than emotions. Nonetheless, moral opinions seem to be closely bound to these more tangible and temporary states. If we take ourselves to have done something morally wrong, we tend to feel guilt, and if we take someone else to have done something morally wrong, we tend to feel anger or indignation: indeed, the mere imagining of morally wrongful action can evoke emotions such as guilt and anger. In this section we shall speculate about what the connection between moral opinion and moral emotion might be, and for that purpose we need a general theory about the place of emotions in our psychology: in satisfying that need we shall borrow much from Antonio Damasio’s recent suggestion in his brilliant book Descartes’ Error.²

Emotions come in many varieties: happiness (including varieties such as euphoria, ecstasy and Schadenfreude), sadness (melancholia, wistfulness), anger (resentment, indignation), fear (guilt), shame and disgust. Plausibly, they are relatively non-voluntary responses to representations of features inside or outside of the body, responses the function of which is to prepare us for actions of certain kinds (by changing such things as muscular tension, heart rate and the quality and speed of cognitive processing) and – often – to signal this readiness.³ Now, “moral” emotions – emotions connected to moral opinions – such as anger, resentment and guilt typically involve changes in facial expressions that do comparatively little by themselves to

¹ Hare’s version of utilitarianism in Moral Thinking, Oxford: Oxford U. P. 1981, works much like that, suggesting a rather determinate account of what it is to be mistaken in one’s moral judgements.
³ Descartes’ Error, ch. 7, esp. p. 139.
prepare one for specific action (unlike other changes in muscular tension and changes in cognitive processing). That suggests that the evolutionary function of moral emotions is, in part, to change the behaviour of others by signalling the kinds of action that one is likely to take. Obviously, anger prepares one for aggressive action, but displays of this readiness might stop others from doing things that are likely to trigger aggressive action. Guilt, on the other hand, belongs to a kind of responses found in various social species (canines and primates\textsuperscript{1}, especially) that prepares the individual for submissive action, and displays of this tendency function to avoid aggressive action from that party, since there is little need for aggression if there is no resistance to one’s will.

Now consider the kinds of behaviour that we typically take to be morally wrong: stealing, lying, cheating, hurting, killing, letting suffer, letting die, deserting family and friends, not sharing as expected\textsuperscript{2} – all acts likely to be followed by aggressive action (which may or may not involve physical violence) from those affected negatively by the act either directly or through sympathy with or dependence of the victim, and therefore likely to be accompanied by guilt or similar emotions to the extent that aggressive behaviour cannot successfully be met by (threat of) force or by flight.

What we have so far is the beginnings of an explanation of why we tend to feel guilt and anger in connection with the actual performance of what we take to be morally wrongful actions. But what do emotions have to do with moral optations? Consider a small child smelling food of her liking. As an inherited reflex or because of learning, her olfactory representation causes a bodily and cognitive response – a craving, or similar –


\textsuperscript{2}Different species of primates, like different human societies, have different expectations about sharing. Anger is evoked when expectations are foiled. Op. cit., ch. 4.
which functions to prepare ingestion and digestion: salivation, producing plans that might take her to the food, and so forth. In so far as the connection between the representation and the ensuing action is functional rather than a freak occurrence, it consists in a strategy-selecting state which might be (depending on the child’s capacity for voluntary action) an optation to eat the food.¹

If Damasio is right this strategy-selecting state works by causing an emotional reaction which is represented in the somatosensory cortex (the part of the brain responsible for our sense of touch, temperature, pain, joint position and visceral state): that representation, in turn, causes an intention to eat the food. But Damasio also points out other functions of emotions. Not only can emotions cause intention and action, they can keep one’s attention focused on certain options and possible consequences of action. Suppose that the child hears a dog bark in a neighbouring room, and that her auditory representation causes fear that prepares her for defensive action such as running away or hiding. This bodily and cognitive response also keeps the dog’s possible appearance in the child’s attention as she moves toward the kitchen from where the smell comes. Moreover, emotional states can serve to keep certain options away from consideration. Suppose that the child now hears a dog bark in the kitchen. Again, the reaction is fear, but this time the fear is connected to the representation of a location involved in her current plan of action. Associating the fear with the kitchen, options that involve coming near that fearful place are withheld from her mind’s attention.

So, the function of emotions goes beyond preparing the body for action by changing heart-rate, breathing and muscle tension. In particular, emotions fulfil the cognitive functions of (a) stimulating the creation of options, (b) keeping options and represen-

¹ Damasio calls these strategy-selecting states “somatic markers”, since they connect certain representations with the somatic (bodily) reactions involved in emotions. See Descartes’ Error, p. 173.
oration of possible events in working attention and (c) keeping options from entering or staying in working attention.

To be sure, this illustration is a schematic picture of a simple example. The child only considers options and possible events that are salient in the situation, but our memory and our capacity to learn allow us to consider all sorts of options and possible events. Moreover, it is plausible that the detour via global bodily and cognitive changes can be bypassed as we mature, let the strategy-selecting state affect choices and deliberation before the entire body is involved and its overall state brought to attention. Damasio’s suggestion is that the strategy-selecting state causes the somatosensory cortex to adjust almost as if the body was in the relevant emotional state, and that it thereby can have some of its usual effects on deliberation.¹ Most of this can take place without being brought to conscious attention, causing us to experience the relevant emotion – to feel happy, proud, angry or guilty. When they do, however, emotions are often experienced in connection to representations that initiated them, thereby allowing us to feel angry with or guilty over the object represented.²

In light of this theory, the following connection between moral optations and moral emotions seems plausible:

Moral optations are formed when attending to moral considerations in practical deliberation. Moral considerations, in turn, tend to affect deliberation by eliciting strategy selecting states the proper function of which is to affect action through moral emotions or their bypass, “as if”, equivalents.

A somewhat similar suggestion is presented by Allan Gibbard in Wise Choices, Apt Feelings. Gibbard can be said to identify moral opinions with optations that have moral emotions as their objects: to have the opinion that one has done something

morally bad is to have the optation to feel guilt under certain circumstances.¹ Now, I do think that most of us are familiar with optations that have emotions as their objects ("I wish I felt differently about this...") but I also think that we can distinguish such optations from moral opinions that have actions as their objects. I might be wrong in this, though, and Gibbard has an argument for assuming a more indirect connection between moral opinion and moral emotions: we can feel guilt while thinking that we have done nothing wrong.² On Gibbard’s account that would be perfectly intelligible since there is little doubt that we can feel guilt while having a certain desire not to.

But the fact that a connection seems roundabout in a few deviant cases does not imply that it is roundabout in standard cases, and a closer look suggests that the simpler model provides a natural explanation of the phenomenon Gibbard aduces. As explained in section 2.4, internalism identifies categorical moral opinions with the results of practical deliberation, and during such deliberation I might want to be influenced by certain aspects but not others. Such second-order optations or wants are not always effective, however, and it might well be that considerations banned from deliberation continue to make their presence known, either as emotions and feelings of emotions or as – in cases of weakness of will³ – ensuing action contrary to judgement.⁴

Of course, if moral optations functions by eliciting emotions, they can be said to be states selecting for emotional responses, which is pretty much what Gibbard suggests that they are.

¹ Wise Choices, Apt Feelings, e. g. pp. 40–5.
³ We return to weakness of will in section 5.3.
⁴ The actions resulting from our moral feelings in spite of our deliberations might be noble as well as ignoble, depending on whether the kinds of consideration that we are willing to bring to bear are sound or not. See Jonathan Bennett’s "The Conscience of Huckleberry Finn", Philosophy, Vol. 49 1974, pp. 123–43, for an interesting reminder of how complex the connection between moral opinion and feelings can be.
Moreover, since (1) one of the functions of anger and indignation over a certain kind of action is to affect the behaviour of others by eliciting guilt in them when they consider or perform that kind of action, and since (2) one of the functions of moral optations is to elicit anger and indignation towards kinds of action, then (3) moral optations can be seen as states selecting for guilt in others upon considering or performing the kind of action that is the object of the moral optation.¹ Again, this seems to be in line with Gibbard’s suggestion. Still, the connection between moral optations and emotional responses in oneself or others is not the right one for moral optations to be seen as optations for emotional responses: practical deliberation leading to a moral opinion is typically not centred on emotional responses but on the explicit object of the moral opinion, namely an action or a kind of action.

Until further arguments are given for a more indirect connection between moral opinions and moral emotions, the suggestion remains that moral emotions (or their bypass equivalents) are the means through which moral opinions / optations affect the actions of oneself and others, not their objects.

2.6. Internalism and the language of morals

In this essay, nothing very informative or sophisticated will be said about the language of morals. Nevertheless, an account of moral opinions will suggest some things about moral language. For example, it is a commonplace that when we utter moral sentences in the indicative, we typically express a moral opinion. (Conversely, moral sentences are perhaps best understood as sentences expressing moral opinions.) Without giving an analysis of what it is to express a particular state of mind, an inter-

¹ David Copp suggests that moral opinions are, roughly, beliefs to the effect that there is a justified standard banning a certain kind of behaviour and calling for regret in those who fail to comply. I believe that a functional analysis of moral emotions captures what is plausible about the latter part of his suggestion. See Morality, Normativity and Society, e.g. pp 25–6.
nalist might adopt and adapt that commonplace, saying that when we utter moral sentences we express moral optations.

Similarly, it seems to be a commonplace that the function of the moral sentence is to communicate the relevant moral opinion from the sender to the receiver, which, under internalism, is to communicate a moral optation from the sender to the receiver. Such communication can be of two kinds: one involving the receiver's sharing the moral opinion / optation of the sender; the other involving the receiver's believing that the sender has the moral opinion / optation. Moreover, in so far as the communicative function of moral sentences is of both these kinds, it is both to make the receiver believe that the sender has the optation expressed by the sentence to and make her have the same optation.\footnote{None of these remarks say much about the nature of communication or the mechanisms of human linguistic communication. Here I defer to a debate to which Dan Sperber and Deirdre Wilson's Relevance: Communication and Cognition, Oxford: Blackwell 1986, is a very interesting recent contribution. For criticism, see commentary and response to commentary on that work in Behavioral and Brain Sciences, Vol. 10, 1987.} If internalism is correct, then, we have a partial vindication of Charles Stevenson's much criticized but rightly famous suggestion that sentences like “This is wrong” should be analysed as meaning, roughly, “I disapprove of this; do so as well”\footnote{See Stevenson's Ethics and Language, especially ch. 2}, where the two parts of the analysans indicate the two communicative functions of “This is wrong”.

2.7. Our Choice Internalism

We have clarified the interpretation of internalism in several ways, and it is now time to sum up. The internalism to be discussed in the ensuing chapters claims, first, that there is a kind of psychological state – a “moral optation” – that is strategy-selecting and formed in practical deliberation concerned with “moral” considerations and, second, that moral opinions are moral optations:
(A) *Opinions* to the effect that performing an act is *morally wrong* are moral optations for the act *not* to be performed.

(B) *Opinions* to the effect that *not* performing an act is morally wrong are moral optations for the act to be performed.

What has been said in this chapter has concerned *assertoric* moral opinions: opinions saying that such and such is wrong or obligatory, period. But we also have *permissive* opinions saying that such and such is neither wrong nor obligatory, as well as *hypothetical* moral opinions to the effect that something is wrong *if* such and such is the case, or that something is wrong *if* something else is right, and so forth. The internalist account of such opinions will not be of paramount importance until section 4.5, however, and we will save it for then.

Our access to moral optations can be quite indirect, depending largely on our theorizing about and models of the functional organization of our action-guiding systems. The fact that I am in a state the proper function of which is to cause certain actions under certain circumstances need not have any upshots whatsoever in overt behaviour unless it is working properly and unless the relevant circumstances obtain. A consequence of this is that moral internalism becomes a theory that must be assessed in light of fairly intricate evidence.

When we understand the debate over internalism as a debate about such things, it makes sense that there is widespread disagreement and that people have different intuitions about whether there can be moral opinions unaccompanied by optations (as noted in section 1.2): the issue need not be just a matter of words as one might otherwise have suspected. And the disagreement makes even better sense when we remember that emotional issues of truth and rationality in ethics have been taken to depend on the truth of internalism. Eventually, we will see how current evidence supports the internalist camp. Before we start examining various arguments for and against
that claim, however, we need to have a reasonably clear picture of what an interesting alternative view would be.

2.8. The externalist alternative

If internalism is correct, a person’s moral opinions consist of optations (desires or wishes) that actions be performed or abstained from. Most of the arguments for and against this claim have the form: “Certain facts are best explained on the assumption that internalism is false (true) and so it should be abandoned (accepted).” In order to assess such claims, it will be necessary to have some alternative story to compare with internalism. Presenting such an alternative – “externalism” – is the object of this section.

Now, “externalism” has often been used as a label for the negation of internalism, but that practice will not be adopted here. The reason is that the negation of internalism is indeterminate – there are many ways in which internalism could be false – too indeterminate to provide any substantial explanatory alternative to internalism. A first criterion for an adequate characterization of externalism, then, is that it should be a competitive alternative to internalism in explanations of our various intuitions and observations concerning moral opinions and motivation. Moreover, since internalism has been seen as a threat to the notion that moral opinions fall into the realm of truth, falsity and rational criticism, an adequate externalist alternative should preserve that idea.

Objectivism The natural way to preserve the idea that moral opinions can be true or false as well as rationally criticized when false is to conceive of moral opinions as representational states, representing or mapping actual and possible actions and their moral status. If this view – call it representationalism – is correct, there is a straightforward sense in which moral opinions can be rationally criticized when performing their function unsuccessfully. And insofar as correct mapping can be
equated with truth, representationalism makes it possible for moral opinions to be true or false.

Representationalism is not enough to preserve what internalism has been supposed to threaten, however. Suppose for example that what we might call “subjectivism” is true: that a person’s opinions about the moral status of his actions represent or map some idiosyncratic fact about his optations: whether he desires to perform these actions, say, or whether he has a desire to (not) desire to perform them. That way, his moral opinions could be true or false, but for reasons limiting rational criticism in ways that would displease our externalist. Intuitively, ignorance of one’s optations is likely to distort one’s moral vision, as this subjectivist theory would explain, but champions of truth and rational criticism in morals are unlikely to accept that it is the only ground for legitimate criticism. Neither would they be willing to give up the idea that one of us must be wrong if I am of the opinion that, say, euthanasia is morally permitted but you are not: supposing that I desire some acts of euthanasia to be performed but you do not, our moral opinions could both be true on the subjectivist account. Furthermore, the relation between moral opinions and optations suggested by subjectivism is too tight to provide an interesting alternative to internalism, since the relevant optations will accompany moral opinions whenever they are correctly represented by the judge.

For such reasons it is better to think of externalism as incorporating not only representationalism but also objectivism, the (not very precise) thesis that the truth of moral opinions does not depend on the idiosyncratic optations of the moral judge. (Objectivism thus includes various forms of relativism, idealism, constructivism and contractualism — as long as they do not make moral properties dependent on factors that are intimately connected to the optational idiosyncrasies of the moral judge.) I know of no alternative view of moral opinions
that can provide the basis for a viable alternative to internalism.¹ So, although I believe that subjectivism is more plausible than many seem to think (provided that it is given a careful interpretation), I will somewhat dogmatically assume that it is false, and keep it out of our discussion for the rest of this essay.

Externalism Although internalism is most strongly associated with attacks on objectivism in recent meta-ethical debate, things have not always been that way and there are still objectivists who think of themselves as internalists.² Moreover, our characterization of internalism makes no reference to the (lack of) objectively representational status of moral opinions, which makes it logically compatible with the idea that moral opinions represent objective facts.

Of course, internalists must hold that moral opinions are optations, too – one state having both representational and strategy-selecting function³ – and this cannot be if Humean anti-rationalism is true: if desires cannot be true or false. But in order to guarantee that internalism and externalism are com-

¹ Note that objectivism does not, at least not in any simple way, imply that there are acts that instantiate moral properties or that we have knowledge of such properties. Objectivism says what the function of our moral opinions is, not that it is actually served.
³ Somewhat similarly, my portable computer keeps track of or represents its battery level, sometimes going into a state the function of which is to represent low battery level and to start a routine that informs the user that the battery level is low, as well as a routine that cuts energy expenditure.

peting theories whether or not some anti-humean view is correct, and in order to take seriously the idea that moral motivation is something external to moral opinions, we shall take externalism to be the view that moral opinions are objective representations that do not consist in optations. Instead of being part of our very moral opinions, externalism takes moral motivation to resemble our motivation to act on beliefs about such things as the effects of our actions on the emotions of our near and dear or on our salaries. Most of us do care about such effects, and to the extent that we do, our beliefs that they follow from certain of our actions will provide motivation. Our care is not part of such beliefs, however: we sometimes have the beliefs without being motivated, and when we are motivated to act according to our moral opinions, such motivation is explained with reference to various desires or tendencies to form desires that often, but not always or necessarily, accompany moral opinion.  

Admittedly, this has been a sparse account of the externalist alternative. But the objectivist thesis will be discussed in chapter 4, and more features of externalist explanations of motivation will emerge in externalist responses to the internalist arguments of the next chapter. The time has come, then, to see which of the two theories we have most reason to believe in.

1 David Brink’s *Moral Realism and the Foundations of Ethics* is a typical representative of the externalism with which we will be concerned.
3. Traditional arguments

To give a convincing verdict on internalism, we need to find some observation or premiss about which internalists and externalists can agree, and show that it supports internalism over externalism, or vice versa: this is what the next three chapters are dedicated too. As it turns out, the arguments that have previously been put forth in the debate are inconclusive. However, since a discussion of arguments for internalism and externalism will not only test the validity of these theories, but also make our conception of them – as well as of morality – richer, this discussion is still worth our while. If nothing else, it will give us an idea of why philosophers have failed to provide generally convincing arguments, and lead us to an argument that shows more promise than the others.

Since we have defined externalism so as to include objectivism one way to argue that internalism is more plausible than externalism is to give reasons to deny that moral opinions are objective representations. But since much of the interest in internalism has been motivated by a belief that the truth of internalism will imply the falsity of objectivism, and since the latter thesis has been at the focus of extensive discussion, we shall start our examination with arguments that avoid assuming any particular outcome of that discussion.

3.1. An argument from relativism

The following passage from Charles Stevenson’s classic on emotivism, *Ethics and Language*, probably captures intuitions that have motivated many internalists.

People with different racial or temperamental characteristics, or from different generations, or from widely separated communities, are likely to disagree more sharply on ethical matters than on factual ones. This is easily accounted for if ethics involves disagreement in attitude; for different temperaments, social needs and group pres-
sures would more directly and urgently lead these people
to have opposed attitudes than it would lead them to
have opposed factual beliefs. The contention that ethics
involves disagreement in attitude thereby gains in prob-
ability. It finds confirmation from what it explains, like
any other hypothesis.¹

Although Stevenson talks about disagreement in attitudes,
rather than in optation or motivation, it is clear that attitudes,
as Stevenson conceives of them, involve optations. An attitude
is “a complicated conjunction of dispositional properties ... marked by stimuli and responses which relate to hindering or
assisting whatever it is that is called the ‘object’ to the atti-
tude”.²

If we disregard the unnecessary reference to “racial charac-
teristics” and substitute factors such as material welfare, social
standing and gender, Stevenson’s observation about moral dis-
agreement seems to put the finger on something that might at
least inspire internalism.³ It is striking that differences in moral
opinions often follow material, social, and psychological factors
that are known to influence motivation, and this indicates a
fairly strong connection between moral opinions and optation.
People with high incomes tend to think that high or progressive
income taxes are unfair to a larger extent than the poor; men,
who generally take a lesser interest in the welfare of others
than women and a greater interest in power and violent conflict,
tend to think that a higher defence / health-care ratio in govern-

¹ Ethics and Language, p. 18.
³ This is not to say that Stevenson himself hoped to establish internalism with
reference to that observation: he wanted to establish something much weaker,
namely that when we try to change each other’s moral or ethical views we (often)
hope to change each other’s attitudes (and motivation). See Ethics and Lan-
guage, ch. 1. Although Stevenson has, on occasion, expressed internalist sympa-
thies, Ethics and Language contains (to my knowledge) no clear explicit internal-
alist claim, being concerned with moral speech and thought rather than moral
opinions.
ment spending is justified;\(^1\) and journalists seem to be morally more agitated than others when freedom of speech is threatened. I am sure that the reader will be able to provide further examples of how people's moral views differ in accordance with differences in features that are likely to affect their optations—call such features \(o\)-features. But to make evident a strong connection of sorts is far from showing that it is internal rather than external, as we will see.

The internalist argument based on Stevenson's observation claims that internalism provides better explanations of the following statement:

\[O\text{-relativity}: \text{Disagreement on moral matters is more likely to follow } o\text{-features than is disagreement on factual matters.}\]

In absence of detailed and empirically well-founded data, the interpretation of \(O\text{-relativity}\) depends on our fairly vague grasp of \(o\)-features, on vague statements about the likelihood of disagreement, and on a border between factual and moral matters that is at least questionable. These difficulties will be left without much discussion, but note that the distinction between the factual and the moral should not be understood as implying that moral matters are non-factual in the sense of being non-representational: the expression "factual matters" signifies factual matters that are non-moral; whether moral matters are factual or not is left open.

\(^1\) My information about the justification of defence versus health-care spending comes from opinion polls presented in Swedish newspapers and television. (Since I have lost the exact reference, it should be taken as no more than an example.) A similar difference between the sexes (with plenty of exceptions, of course) seems to exist for very small children, though. According to a survey of research on differences between the sexes, baby girls are more responsive to social clues and more prone to empathy from birth than baby boys. See Anne Moir and David Jessel, *Brainsex: the Real Difference Between Men and Women*, London: Joseph 1989. The book starts out as a sober survey of research but ends on a considerably more speculative note. The present point seems well founded, however.
Establishing o-relativity  The first task for the internalist is to establish o-relativity, something which offers prodigious problems. The externalist might point to the well-known fact that in non-moral questions of great practical importance – questions concerning the effects of certain political decisions, for example – our judgements tend to be influenced by our optations. Since we regard moral questions as being of great practical importance, we should compare disagreements in moral opinion with disagreement in non-moral factual matters that are of great practical importance. It is not obvious what such a comparison would show, especially if we hold in mind that it is when matters of great practical importance are epistemically evasive that our optations, hopes and fears have most leeway: the externalist might well insist that moral opinions are based on beliefs about factual matters that can be extremely hard to ascertain. (As anyone with an open mind and experience of philosophical ethics knows, it is often very difficult to know what is right and what is wrong – at least in cases where there is much controversy; the ones with which this argument is concerned.) Questions about the moral status of an act seem to depend upon questions concerning intentions and feelings, as well as consequences of the action judged and its alternatives, and such questions are seldom open to straightforward decisions based on easily available facts. If we are to compare moral opinions with paradigmatically externalistic beliefs, the latter must concern questions that are equally hard to answer, depending on equally complicated and evasive features of the world, and of equal practical importance. It is not clear that a comparison with such beliefs would show that disagreement on moral matters is more likely to follow o-features than is disagreement on factual issues.

Explaining o-relativity  If the internalist had means to establish o-relativity, he could say that disagreement on moral matters follows differences in optation more than ordinary factual
opinions because the opinions on moral matters are options, which the latter are not.

As we saw in the previous section, however, o-relativity will only be tenable if we do not qualify the comparison between the moral and the factual case as being one between moral opinions and a subclass of comparable factual opinions: opinions that concern practically interesting features of the world and are epistemically evasive. But an unqualified version of o-relativity, although more tenable, would lack relevance, since the qualifications would be available for the externalist in the explanation of that version. He might say that even though there might, in general, be more disagreement following differences in o-features in moral matters than in factual matters, this is only because moral matters tend to be epistemically more complicated and practically more important than factual matters in general.

The externalist response to the argument from o-relativity has invoked auxiliary hypotheses concerning the content of moral opinions. It was claimed that the assumptions on which we base our moral opinions – our moral reasons – tend to be epistemically intractable; hard to verify or falsify because of the tricky psychological and counterfactual issues involved. It was also claimed that moral wrongness and rightness are properties of great practical importance. Since these are reasonable hypotheses given that moral objectivism is true (which we will assume until chapter 4), an argument for internalism departing from o-relativity faces enormous problems, and we shall direct our search elsewhere for the remainder of this chapter. It will become evident, however, that most of the traditional arguments for internalism are faced with similar problems.

3.2. More signs of the practical nature of morality
Other observations pertaining to the practical nature of morality have been invoked in support of internalism. Here is Stevenson, again:
Suppose that we are trying to convince a man that something he did was wrong. He replies: "I fully agree that it was, and for that very reason I am all the more in favor of doing it over again." Temporarily puzzled to understand him, we shall be likely to conclude, "This is his paradoxical way of abusing what he considers our outworn moral conventions. He means to say that it is really right to do it, and that one ought to do it flagrantly in order to discredit the many people who consider it wrong." But whatever we may make of his meaning (and there are several interpretations possible) we shall scarcely take seriously his protestations of agreement. Were we not trying all along to make him disapprove of his action? Would not his ethical agreement with us require that he share our disfavor – that he agree with us in attitude? ¹

Stevenson's last rhetorical question suggests that he hoped to have elicited intuitions in favour of internalism, and we can extract from our quotation a number of observations the likes of which have subsequently been used as premisses in arguments for internalism. For example:

(a) When someone makes an assertion about what she morally ought or ought not to do, we expect her to have the relevant optation; assertion of a moral sentence (concerning one's own behaviour) is surrounded by an air of paradox when coupled with indifference.

(b) Lack of moral optation is taken as a sign of insincere assertion, or lack of moral belief.

(c) Our standard of success in moral argument is not a cool, indifferent agreement, but rather the adoption of certain attitudes and behaviours. Were we to set out to

¹ *Ethics and Language*, pp. 16–7.
persuade the sexist of the evils of his practice, we should feel that we had succeeded only if we noticed some change in his behaviour toward women, or at least some motivation to behave otherwise.\(^1\)

Again, the idea is that internalism can provide better explanations of signs of the practical nature of moral discourse and thinking than externalism.

**Internalist explanations** Indeed, the observations above seem readily explained if we assume that moral opinions (and – consequently – sincere moral assertions) are connected with optations in the way suggested by internalism. Take (a). We typically assume a moral assertion to express the speaker’s moral opinion – only then is indifference paradoxical in the relevant way. From that assumption, taken in conjunction with internalism, it follows that the speaker has the relevant optation. Similarly, internalism holds that lack of moral optation implies lack of corresponding moral opinion, which makes the expression of that opinion insincere. Thus the connection between sincerity and optation claimed in (b) is explained. Finally, (c) is explained by the fact that if we want the sexist – or whoever it is that we want to convince or convert – to see that his behaviour is morally wrong, we obviously want him to have the opinion that it is morally wrong, and internalism claims that coming to have that opinion involves coming to be motivated to act differently. But as with the preceding arguments for internalism, the externalist seems to be able to provide plausible alternative explanations.

\(^1\) The observations (a), (b) and (c) are taken more or less verbatim from M. B. E. Smith’s “Indifference and Moral Acceptance” in *American Philosophical Quarterly* Vol. 9, No. 1 1972, p. 89. Appeal to (a) is still in vogue, see for example Jonathan Dancy’s *Moral Reasons*, p. 4, and Michael Smith’s *The Moral Problem*, e.g. p. 6.
Externalist explanations  On externalism it is a contingent matter whether someone who has a moral opinion is motivated to act accordingly: most are, but some are not. Still, this seems to provide externalism with the relevant explanatory resources. Take (a). Why are we perplexed by people who are completely indifferent to what they say is their moral duty, or to what they say they ought, morally, to do? The best externalist answer is that when we describe something as our moral duty, or as what it would be morally wrong not to do, we ascribe a property to that action which is such that very few would be indifferent to whether their actions exemplify that property or not. We particularly like others to perform actions exemplifying it, and we strive ourselves to perform such actions. It is a sad fact that not everyone shares this striving, but the few exempted from it by Mother Nature and Father Culture can often discern whether actions exemplify such a property or not, and have moral opinions – opinions ascribing that property to some action – without being moved to act. The perplexity alluded to in (a) concerns that person: how can he be indifferent to that action if he believes those things about it; what kind of person can he be? Does he really believe that the action exemplifies the property in question?

Observation (b) is explained similarly. Because most people care, more or less, about not doing anything morally wrong we do expect them to be motivated according to their opinions, and our expectation is further supported by the fact that the speaker made the moral assertion. In many cases it would be hard to see why anyone would have formed moral opinions – let alone given them voice – unless she was interested in acting according to them: the very utterance raises the (initially high) probability that the speaker cares about whether her actions are morally wrong.

The externalist explanation of (c) is that we care about people doing what is morally right and refraining from what is morally wrong. We often engage in moral discussion in order to make the other person see that some behaviour of his is morally
wrong by pointing out in what ways it is morally wrong, and if we suspect that he cares little about them, we try to instil the relevant feelings by means of telling examples and apt rhetoric. But not always. At academic seminars, for example, people sometimes engage in moral argument out of curiosity or just for the fun of it. On such occasions they might be satisfied with cool, indifferent agreement. Even on such occasions, though, lack of motivation might indicate insincerity — as noted in (b) — which, in turn, indicates that the professed agreement was only apparent.

Concluding remarks We have now seen internalist and externalist explanations of various signs of the practical nature of morality. Unfortunately, neither set of explanations wears “explanatorily superior” on its sleeve. The reason is that although externalism denies that moral opinions are optations, this is consonant with a strong statistic correlation between moral opinion and corresponding motivation — and statistic correlations are enough to ground strong expectations of the kind that are needed to explain (a), (b) and (c). Again, more sophisticated arguments are needed to decide the issue.

3.3. Why does “ought” imply “can”? One thing about our moral opinions that might support internalism is their sensitivity to considerations concerning what we can or cannot do: unless my thinking is seriously muddled, I will give up my opinion that I ought to rescue a person when I find that she cannot be rescued. If it can be shown that internalist explanations of this fact are superior to externalist alternatives, we have support for internalism. (An argument along such lines is proposed by Richard Hare in *Freedom and Reason*.1)

1 Oxford: Oxford U. P. 1963, ch. 4. The argument, presented on p. 51, goes as follows:
Two incomplete explanations  An internalist explanation should start from the observation that you cannot (rationally) be motivated to perform or not to perform some action if you fully believe that it cannot be performed anyhow. (This is not to say that you cannot rationally be motivated to try performing it even if you believe that you are most likely to fail, but even trying seems to imply having some hope of success, however small.) From this assumption about rational motivation and the internalist assumption that moral opinions involve motivation, it follows that you cannot rationally have moral opinions that involve motivation (not) to do something you fully believe to be impossible. That, in turn, would seem to explain why we are impressed by considerations of performability, at least when thinking clearly.

The externalist, on the other hand, will refer to the nature of moral rightness and moral wrongness and to the externalist assumption that moral opinions represent moral rightness and wrongness of actions. It is part of the nature of properties such as moral rightness and moral wrongness that if an action cannot be performed it is neither morally wrong nor morally right. To fail to grasp this is to have misunderstood what moral

It is because I can act in this way or that, that I ask, “Shall I act in this way or that?”; and it is, typically, in my deliberations about this “Shall I?” question that I ask the further, but related question, “Ought I to do this or that?” Thus it is because they are prescriptive that moral words possess that property which is summed up, perhaps over-crudely, in the slogan, “Ought’ implies ‘can’.” If descriptivism were a complete account of evaluative language, this slogan would never have arisen.

When Hare says that moral words are prescriptive, he is claiming that the acceptance of judgements of the form “X ought to do Y”, when used as value-judgements in Hare’s sense, logically implies acceptance of the prescription “Let X do Y.” (The Language of Morals, p 168–9.) In the case where I am X, the acceptance of “Let X do Y” implies a decision on my part to do Y, and that, in turn, implies that I am sufficiently motivated to do Y. In other words, prescriptivism implies a form of internalism where the relevant motivation is restricted to actions that can rationally be prescribed. Here, however, we will avoid the peculiarities of the prescriptivist doctrine and ask whether Hare’s observation can support the internalism of chapter 2.
rightness and wrongness is, or perhaps to have misunderstood the meaning of the expressions “morally wrong” and “morally right”.

The internalist might complain that this is little more than a statement of the fact to be explained, couched in terms agreeable to the externalist: the real question is why expressions such as “morally wrong” cannot literally and rationally be applied to actions that cannot be performed, and here the connection to motivation and action seems too obvious a candidate not to be true. But the internalist’s explanation is equally incomplete, since it tells us only why moral opinions about one’s own present action are impressed by considerations of performability. Whereas it does seem clear that it is irrational (or perhaps impossible) to be motivated to perform actions that are impossible to perform by one’s own lights (because they are the actions of others or of times past) it is less clear that we are irrational if we wish for such actions to be (or have been) performed: such wishes do not in any obvious way lead us to consciously embark on actions that cannot reach their destination. Reference to the optational nature of moral opinions does not, in itself, explain why moral opinions are sensitive to information about performability in fairly determinate ways. Both the externalist and the internalist responses need amendment.

A piece of armchair etymology Why do we employ concepts of moral rightness and wrongness that imply performability? No doubt because such concepts are practically interesting in a fairly direct way. They are used in personal planning, in the rearing of children, and in the coordination of complex social life in a way that they could not be if we did not assume them to concern performable action. If forming moral opinions about actions is to guide optation and action, it must be affected by considerations of performability or lead to frequent disasters for their agent. And if voicing one’s moral opinion is to be one way of advising or commanding, it would be much better if we could assume that the action recommended or commanded on its
moral merit (on moral grounds) was performable, or at least believed to be performable.

This account might seem to be in the spirit of the failed internalist answer, but it provides an explanation of a missing premiss: moral opinions are connected to optations that are especially likely to have practical upshots, namely the optations concerning performable actions that lead us to recommend or command such action. Still, to explain why it seems to be a conceptual truth that moral wrongness (and rightness) implies performability there is no obvious need to assume that all opinions and all sincere discourse about moral wrongness are intimately connected to action: concepts are presumably formed and upheld in response to fairly constant needs for communication and information processing, but these needs are not necessarily served every or even most of the time the concept is employed. Enough is enough.

So, since internalists and externalists alike are entitled to the assumption that our moral concepts are practically interesting and that they would be much less so if they did not imply performability, they are equally capable of explaining why wrongness implies performability, even if the externalist denies the internalist connection between moral opinion and optation.1 The internalist is right to insist that it is the connection between moral opinion and action that explains the connection between moral opinion and performability, but further arguments are needed to establish that moral opinions are con-

---

1 For those who think that the moral status of an action depends on the value of its consequences as compared to the value of the consequences of alternative actions, skipping the demand for performability entails a major problem. The problem is to identify the alternatives relevant for comparison without recourse to a notion of performability: the natural thing to say is, roughly, that the relevant alternatives to an action are the actions that could have been, or can be, performed instead. For problems concerning that rough conception of alternative in consequentialist ethics, see Lars Bergström’s The Alternatives and Consequences of Actions, Almqvist & Wiksell 1966. For a defence, see Erik Carlson’s Consequentialism Reconsidered, Dordrecht: Kluwer 1995.
nected to action by being constituted rather than frequently accompanied by optations.

3.4. Hare’s argument from practicality

Internalists and externalists alike should accept that moral opinions and their expressions play important roles in regulating social conduct and personal action. Utterances of sentences like “You must return his calls” are frequently taken as answers to practical questions, questions concerning what one shall do. Similarly, the thought that it would be morally wrong to act in a certain way will often rule that act out of our practical deliberation. Such (trivial) facts about moral speech and thought form the basis for the following internalist argument presented by Richard Hare:

A statement ... cannot answer a question of the form “What shall I do?”; only a command can do this. Therefore, if we insist that moral judgements are nothing but ... statements of fact, we preclude them from fulfilling their main function; for their main function is to regulate conduct, and they can do this only if they are interpreted in such a way as to have imperative or prescriptive force.1

Since the argument is stated in technical terms that may be unfamiliar to the reader, we shall start with some clarifications.

Interpreting the argument The phrase “moral judgements”, as used by Hare, typically designates utterances or speech-acts (including, perhaps, mental acts such as thoughts couched in sentences) rather than states of mind or opinions. The argument is thus not directly concerned with moral opinions.

“Command” is a technical term meant to cover “all these sorts of thing that sentences in the imperative mood express”,

1 Freedom and Reason, p. 46.
for example “military orders ... architects’ specifications, instructions for cooking omelets or operating vacuum cleaners, pieces of advice, requests, entreaties, and countless other ...”.1 Perhaps we could say that to command is to make an utterance the communicative point of which is to get someone to do something. It is not the form of the sentence uttered that matters: “The best thing you could do is to scram” displays the indicative form but is likely to be uttered as a command — in an effort to tell someone to scram. Without going into difficulties with the analysis of commands, it seems obvious that sincerely making a command involves wanting the command to be obeyed and that accepting a command for one’s own present action involves having the intention to perform the action in question.2

If Hare is right in his claim that making a moral judgement is (or implies) the acceptance of a command, making a moral judgement will imply having the relevant optation, since having the intention to perform an act or wanting someone else to perform it involves having the optation that it be performed. And if making a moral judgement implies having a certain optation then the state of mind that we express when we make moral judgements about our own present case seems to involve that optation: and that state of mind is plausibly seen as a moral opinion. If Hare is right, then, we seem to have an argument for internalism that starts from the idea that moral judgements are answers to practical questions.

The argument starts from premisses, (F) and (C), from which the conclusion, (J), or, in extension, (O), is supposed to follow:

(F) The main function of moral judgements is to answer questions of the form “What shall I do?”.

(C) Only commands can answer questions of the form “What shall I do?”.

1 The Language of Morals, p. 4.
ergo

(J) Moral judgements are or imply commands.

(O) Moral opinions are, in part, desires.

It is not at all clear, though, how the premisses are supposed to establish the conclusions. Before looking at the problems of the argument, however, we need to get the interpretation of the counter-intuitive (C) right.

In one sense, a practical question of mine has been answered when I have received information or impulses that end my query, by making me decide what to do. If I am trying to decide whether to have fish or soup for lunch (“Which shall I have?”), the smell of divine soup from the table next to mine, or my friend’s question about what I had for lunch yesterday could effectively answer my question in this sense. If we interpret (C) along this line, it is of course an embarrassing blunder – especially since Hare makes no effort to substantiate this premiss.

But in another sense, to answer or give an answer to a practical question such as “What shall I do?” is to provide information or impulses the point of which is to make the person posing the question choose in a particular way; in other words: to command.1 That is how my friend’s utterance of the sentence

1 In his “A taxonomy of illocutionary acts”, reprinted in Expression and Meaning, Cambridge: Cambridge U. P. 1979, John R. Searle uses the term “directive” for (approximately) the kind of speech acts referred to here as “commands”. Searle classifies speech acts in terms of their illocutionary point, that is, the function for which they are used, and “The illocutionary point of [directives] consists in the fact that they are attempts by the speaker to get the hearer to do something.” (p. 13) If moral judgements are directives, a sincere moral judgements will involve sincerely wanting the directive to be followed.

To make an utterance the point of which is to get a person to do something without it being manifest to that person that this is the point is to manipulate, not to command. Exactly how to characterize the point of an utterance in the right way, or how to understand what it is for a point to be manifest to the addressee, is a matter of some controversy. For problems with characterizing acts of communication in terms of intentions as Paul Grice does in “Meaning”, Philosophical Review Vol. 66, 1957, and in “Utterer’s Meaning and Intentions”, Philosophical Review Vol. 78, 1969, and as Dan Sperber and Deirdre Wilson do
“Have the soup!” is an answer to my question. Interpreted along that line (C) becomes trivially true: if something is an answer to a practical question it must be a command. That is more likely to be the interpretation that Hare intended, and it is the one that we will assume as follows.

Problems with the argument

Given the trivial interpretation sketched above, we should all grant the truth of (C), and although (F) needs clarification it has an air of truth around it. But as we will see, (F) must be given an interpretation that begs the question against the externalist if (J) is to follow.

In one sense, the main function of an entity is the most important thing that it actually does, and if moral judgements actually answer practical questions, they will be commands. But the idea that all moral judgements are answers to practical questions seems much less plausible than the internalist thesis and is something that externalists could deny with impunity. Sometimes when I say that it would be wrong of me to do something, I say it in order to explain a decision of mine, without any intention to keep myself from doing it: I have already made the decision. (Of course, there are cases where we tell others about our moral opinions in order to make ourselves stick to decisions, but these are not the only cases — far from it.) At other times the purpose of making a moral judgement is to exemplify a good (or bad) argumentative pattern in order to assure someone that an argument (which need not concern her or my actions) is valid (or invalid). Without further arguments, and I can see none in the offing, this interpretation of (F) will be of little use as a premiss in Hare’s argument.

—in their Relevance, see Ruth Garrett Millikan’s Language, Thought, and Other Biological Categories, ch. 3, and her “What Peter Thinks when He Hears Mary Speak” (commentary on Sperber and Wilson’s Relevance, in Behavioral and Brain Sciences Vol. 10, 1987) as well as Simon Blackburn’s Spreading the Word, Oxford U. P. 1984, ch. 4. For a defence, see Sperber and Wilson’s response to commentary on their Relevance in Behavioral and Brain Sciences Vol. 10, 1987.
In another sense, the main function of a thing is what it does most often. Although that will yield a more plausible interpretation of (F) — that moral judgements most often answer questions of the form “What shall I do?” — it will not support (J) since it is compatible with an externalist interpretation of that phenomenon: when moral judgements answer practical questions they are accompanied by an external desire that the answer/command is followed.

In a third sense, the main function of a thing is what it is created or picked out to do by decision or by natural selection rather than what it actually or most often does. Call this its normative function. We might say that the normative function of the sperm is to fertilize an egg — it is because sperms have fertilised eggs that genes for sperm-production have been selected — and we might say that the function of a pair of scissors is to cut. Now, in some sense it is surely true that moral judgements have been selected to guide action — if they did not serve that function, we would not take such an interest in making them. Since commands are utterances the point of which is to guide action, could this show that moral judgements are commands? Not in any straightforward way, and particularly not if that is supposed to show that moral opinions are desires. Consider the following analogy. Unless utterances such as “There are peaches over there” sometimes resulted in the eating of peaches found at the intended location, we would probably cease to make them: whatever other interests we have in peaches are mostly the consequences of peoples’ interest in eating them. But even if that would make us say (oddly) that the utterances in question are answers to practical questions and hence commands it would be downright absurd to say that my belief or opinion that there are peaches at a certain location is a wish or desire that these peaches be eaten. What the externalist should say, then, is that morally good or bad actions are like peaches in being practically interesting in a certain way — some of us are interested in eating peaches; many of us are interested in our own and others’ acting rightly — but that these interests do not
constitute our beliefs concerning the location of peaches and the moral qualities of actions: they are merely their frequent company. An internalist trying to make Hare's argument hold with reference to the normative function of moral judgements must explain how this externalist answer goes wrong.

Since I find no interpretation of (F) that is capable of yielding (J) and (O) without the addition of some extra non-obvious premisses that Hare has failed to supply, I conclude that the argument fails to establish internalism. What this and the previous arguments remind us of is that it is when moral opinions are connected to desires that we take most interest in them.¹ That goes a long way to show that moral opinions accompanied by desires are an interesting kind of judgements for those trying to understand morality, but it does not show that the internalist conception of moral opinions provides a better understanding of morality than externalism.² Some other argument is needed.

¹ To be fair to the spirit of Hare's writing, we could say that he is right in claiming that when moral judgements (considered as speech-acts) are put to (one of) their most important use(s), they are commands, and so accompanied by motivation. Indeed, Hare makes it a matter of pragmatically motivated stipulation that the kinds of judgements that concern him are commands. See The Language of Morals, pp. 168–9.

² Unless we had moral optations, it is very unlikely that we would have the concepts of right and wrong: such concepts would have no use in our psyches. If this is right, externalists and internalists alike should take the development of moral optations as a central part of the natural history of morality. Why, then, have moral optations emerged? First, note that many of our reasons for moral opinions concern relations between people: they are answers to the question: “Who did (not) do what to whom?” Other moral reasons concern the consequences of our acts upon the welfare of others, rather than just ourselves. Between the two, these kinds of considerations presumably cover almost every belief that we readily accept as a moral ground. Our question about the history of morality, then, is a question of why we have come to form optations elicited by such moral considerations: why have we come to put practical weight on the details of relations between people, and why have we come to care – to some extent – about the welfare of others?

The answer is most probably an old one: human beings are social beings, and social beings take interest in their relations to – and the welfare of – others, an interest connected to feelings of guilt, shame, resentment and anger. From
3.5. Smith’s argument for internalism

In all arguments discussed above, the presence of moral optation has done the heavy explanatory work. The problem has been that externalism is entirely consistent with the frequent and frequently reliable presence of moral optation: what it denies is that optations are internal to or constitute moral opinions. To vindicate internalism, it must be shown that the tighter connection between opinion and motivation is needed in some other explanations – and that is exactly what Michael Smith purports to do in his recent book *The Moral Problem*. The argument runs as follows:1

By all accounts, it is a striking fact about moral motivation that a change in motivation follows reliably upon a change in moral [opinion], at least in the good and strong-willed person. A plausible theory of moral [opinion] must therefore explain this striking fact. As I see it, those who accept [internalism] can, whereas the ...
externalist cannot, explain this striking fact in a plausible way.

The argument is another piece of inference to the best explanation, where one hypothesis is defended against competing hypotheses on the ground that it explains some relevant fact(s) better than they do. In the present case, Smith argues that internalism explains the following (alleged) fact better than does externalism:

*Reliability:* For the good and strong-willed person, a change in motivation follows reliably upon a change in moral opinion concerning actions open to her.

Smith hopes that externalists and internalists alike can subscribe to reliability; that it is a “striking fact” to be accounted for. Perhaps he is right on that score: reliability is not obviously in conflict with externalism since the externalist can point to the assumption that the person is good in explaining why her motivation is reliably tied to her moral opinions. A good person is one who cares about doing what is morally right and avoids what is morally wrong, and so it is no surprise that a change in view concerning what is right will change her motivation. Hopefully, the fact that reliability seems compatible with externalism allows the externalist’s intuitions about this case to be relatively untainted by his theoretical commitment. (If not, we would still be missing the sought for arbiter between internalism and externalism.)

The internalist account of reliability is quite obvious. According to internalism, moral opinions are essentially tied to specific kinds of motivation, and a change in motivation is an obvious consequence of a change in opinion. If Emily used to have the opinion that it would be morally wrong of her to postpone an appointment, but now has the opinion that it would be morally wrong not to postpone, it follows from internalism that while she is now motivated to postpone, she was motivated not to, which means that there has been a change in motivation. So,
both externalists and internalists can produce an explanation of reliability.

A problem for externalism According to Smith, however, the externalist explanation is flawed in its assumptions about the good person.\(^1\) To be sure, a good person does care about avoiding what is wrong and doing what is right, as the explanation says. But such care can be of different kinds, and the kind of care presupposed by the externalist explanation is not the kind that seems to be essential to being a good person. In particular, mistaken assumptions have to be dragged in when the externalist is to explain the reliable correlation of motivational change and change in fundamental values, or so Smith argues. His example is this: I am planning to vote for the libertarian party, but you convince me that I should, instead, vote for the social democrats, not just because the social democrats will better promote the values I thought would be promoted by the libertarians, but rather because the values I thought should and would be promoted by the libertarians are fundamentally mistaken. You get me to change my most fundamental values. In this sort of situation, what happens to my motivation?\(^2\)

Smith’s answer is that if I am a good and strong-willed person, I will be motivated to vote for the social democrats, just as reliability says.

Before we look at what difficulties this example might pose for externalism, we shall distinguish two interpretations of “fundamental values”. On one interpretation, a person’s fundamental values are the things she cares about the most, the things that, at bottom, makes her tick. But then it would be obvious, regardless of internalist or externalist explanations,

\(^1\) *The Moral Problem*, pp. 73–5.
that a change in fundamental values comes with a change in motivation: it is a change in motivation. A more useful interpretation of Smith’s argument takes it to concern a change in the fundamental principles that, at bottom, determines what her moral reasons are and, in extension, determine her moral opinions. Consider a utilitarian, who takes the consequences of an action and its alternatives upon human welfare to be the one and only thing that determines whether the action is morally right or wrong. In practice, she accepts other reasons for thinking that an action has a certain moral status (that it “feels” right, that it is honest, that it accords with desert) but she takes these to be reasons only so long as she thinks that they are evidence pertaining to the overall consequences for human welfare. We could say about this person that her fundamental (moral) value is overall human welfare.

Some philosophers – often referred to as “particularists” – would argue that what determines a (good) person’s moral reasons is beyond codification; more a mode of vision than the application of principles to particular cases.¹ On their view, our utilitarian represents a degenerate case. To be neutral on that issue, we could say that a change in fundamental moral values is a change in “moral outlook” where this outlook can be the tendency to determine moral issues by applying a certain principle or, alternatively, an uncodifiable way of taking certain characteristics of actions as reasons for one’s moral opinions.²


² For an interesting perspective on the uncodifiability of visual properties that might be applicable to discussions of particularism, see Daniel Dennett’s Consciousness Explained, London: Allen Lane 1991, pp. 375–383. Dennett suggests that colours and colour-vision emerged together as a result of some things needing to be seen (a fruit, say) and others to see them (any animal for which fruits are nutritious): colour and colour-vision were thus designed by evolution to fit each other. Dennett compares with another case in which one thing is designed to fit another: tear a piece of cardboard in two, give one half each to
With that said, here is how the externalist explanation becomes problematic, according to Smith: If we suppose that being good involves caring about libertarian values, the assumption that I am good could explain why I was motivated to vote for the libertarians, but then the assumption that I am good after the change in moral judgement is in conflict with the assumption that I stopped caring about libertarian values. If, on the other hand, we supposed that being good involves having the kind of motivation involved in explaining why, after my change in view, I was motivated to vote for the social democrats, then the assumption that I was good before the change is in conflict with the assumption that I wasn’t motivated to vote for the social democrats before the change.¹

Perhaps this is a realistic example, and perhaps the point can be generalized: goodness cannot be identified with caring about certain fundamental values — if it could, changes in fundamental values would be impossible in the good and strong-willed person, which it seemingly is not.² According to Smith that undermines the externalist explanation, because

... commonsense tells us that if good people judge it right to be honest, or right to care for their children and friends and fellows, or right for people to get what they deserve, then they care non-derivatively about these things. Good people care non-derivatively about honesty, the weal and woe of their children and friends, the well-being of their fellows, people getting what they deserve, justice, equal-

¹ The Moral Problem, pp. 73–4.
² That goodness cannot be identified with some determinate fundamental cares is of course compatible with the plausible idea that goodness consists of some disjunction of such cares.
ity, and the like, not just one thing: doing what they believe to be right, where this is read \textit{de dicto} and not \textit{de re}. Indeed, commonsense tells us that being so motivated is a fetish or moral vice, not the one and only moral virtue.\footnote{\textit{The Moral Problem}, p. 75.}

The allegation, of course, is that the externalist view presupposes a picture of the good person which allows only derivative care about such things as honesty, justice, and the weal and woe of children and friends and non-derivative, \textit{de dicto} care only for doing what one believes is right. Externalist goodness becomes moral fetishism, a mere caricature of real goodness.

Exactly how the externalist explanation presupposes that the good person is good in caring about this “one thing”, or how the difference between non-derivative and derivative care, or between \textit{de re} and \textit{de dicto} readings of “what they believe to be right” should be understood is left to be explained. Since Smith’s statement and explanation of the argument are very brief and somewhat lacking in transparency, we will spend some time clarifying an underlying assumption about the workings of our response selecting systems.

\textit{Dispositions to form desires} One way to acquire new desires is to acquire desires for means to (or instances of) something (or some kind of things) that one already desires. But there are other ways that must be involved in our transformation from babies caring only about getting our basic needs satisfied into complex adults with a number of projects of different kinds, projects that frequently conflict with our basic needs.

One possibility is that derived desires turn into non-derived desires as time passes by. If lots of activities, thinking and planning revolve around what started out as a derived goal, we might come to feel that this goal was meaningful in itself, whether it fulfilled other goals or not. We can also think of more instantaneous ways in which new non-derived desires emerge.
Smith suggests that if I desire a number of things and realize that they all have something in common, I might acquire a general desire for things of that kind.\(^1\) Another and more obvious example of (almost) instantaneous creation of desires is provided by love at first sight. Before our Romeo and Juliet met, they had no desire to live and die by each other's side. It is true—trivially true—that they were disposed to form such a desire, but it seems implausible to insist that this disposition is itself a desire in the relevant sense. Yet another example is provided when television broadcasts from catastrophe areas feature children at the mercy of flies, starvation and disease: such programmes sometimes cause strong and apparently non-derivative desires to help these children. Most of those affected by the programmes surely had some general desire to help the starving, and the children on the screen do instantiate starvation: that would account for the emergence of a desire to help the children on television, a desire derived from a pre-existing desire. But the new desire is often much stronger than the general desire, and very much stronger than would be motivated by its mere instantiation of the general desire: most or all of its strength was instilled by watching the programme (which might have increased the strength of the general desire, too). Furthermore, the disposition to form desires in response to vivid experiences of human suffering is not plausibly described as a desire.

Desires in the sense relevant here are things which more or less consciously enter a person's thinking and planning, and which can be affected by such deliberation. There is no reason to suppose that every case of love at first sight is preceded by such a relatively conscious desire to live one's life with someone of the right kind as a goal around which one's thoughts centred—at least not as a goal as important as the company of the

loved one – and no reason to think that everyone inspired to charity by a television picture of starving children already desired the eradication or diminution of starvation as strongly as she now desires the welfare of the children on the screen.

What Smith’s argument shows Having first-hand experience of material and social misery, say, might convince me that the social democratic values of material equality and the feeling of safety from misery associated with the welfare state are more important than the libertarian values of freedom from infringement (as laid out e. g. in Robert Nozick’s Anarchy, State, and Utopia) and at the same time cause a corresponding change in motivation. Even if I have a general desire to do what is right that explains some of my interest in moral questions and efforts to examine and improve my moral opinions, that general desire might come nowhere near explaining the strength of my new conviction: my new interest in equality and improved material conditions for the poor might be considerably stronger than my pre-existing desire to do what is right.

As Smith suggests, something similar might be true if a change in fundamental values is instigated by argument rather than first-hand experience of the objects of evaluation. Realizing that my moral opinions are inconsistent, or less coherent than some accessible alternative, I might change my moral views and, at the same time, change my non-derivative goals. If you convince me that the correct way to form moral opinions is by impartial consideration of the interests and possible experiences of everyone involved and I believe that such considerations would favour the promotion of social democratic rather than libertarian values, your argument might lead me to change my fundamental values, as Smith suggests, and could very well (reliably, the internalist would say) change my non-derivative desires. My new motivation to promote social democratic values

---

1 New York: Basic Books 1974
could not plausibly be seen as a desire for a means to satisfying my desire to be good, or my desire to be consistent and coherent.

It is perhaps conceivable that someone is completely driven by a desire to do the right thing, but – and if this is Smith’s point he is quite right – that is not the kind of person we think of when we imagine a good person. Such a person would indeed be a moral fetishist. So if what we have said is right, the dynamics of moral motivation cannot be correctly understood if moral motivation is thought of as an external desire to do what is right, and if externalism presupposed such an understanding, it would be wrongheaded. But it is easy to think of versions of externalism that explains moral motivation in other terms.

**Why externalism survives**  
As it happens, Smith’s own theory of moral opinions represents a species of externalism (in our terminology, that is) which has the tools to stay clear of the problem posed by his argument. According to Smith, moral opinions are opinions about what we have the best reasons to do or desire in the situation, or what we would want ourselves to do or desire if we were fully rational and had relevant knowledge.\(^1\) On that theory, it is natural to think of a person’s fundamental values as things that she believes that she has the best reasons to promote for their own sake rather than as means to something else, or, in Smith’s vocabulary, what non-derivative goals she believes she has the best reason to have. Let us translate the voting example into terms explicitly fitting Smith’s theory.

I used to believe that I had good reasons to vote for the libertarian party, but you convince me that there are better reasons to vote for the social democrats. What changed my mind was not that I came to think that the

---

\(^1\) *The Moral Problem*, pp. 182–4. Smith assumes that such opinions are objective representations.
social democrats would better promote the values I thought would be best promoted by the libertarians, and which I used to think that I had good reasons to promote for their own sake. Rather, I came to believe that I have worse reasons to non-derivatively desire libertarian than social democratic values. Being a rational person, my non-derivative desires are changed, and I am now motivated to vote for the social democrats.

My new desire for social democratic values is presumably caused by my belief that it would be rational to desire social democratic values, but what guarantees the change on Smith’s view is the fact (let us suppose) that I am a rational person, a person whose desires are responsive to criticism pertaining to the coherence of my desires and the truth of my beliefs; criticism relevant to beliefs about what is most rational to do.¹ This responsiveness is not plausibly seen as a desire to be rational, but rather as an involuntary disposition to form desires. And being responsive in that way is not fetishistic.

Now, I believe that an externalist can produce a more plausible explanation of reliability than Smith’s. Smith’s theory makes motivation dependent on the rationality of the moral judge, but in explaining reliability in terms of involuntary responsiveness it seems more natural to describe the responsiveness in question as goodness and strength of will (as in Smith’s own formulation of reliability) than as rationality:

¹ Smith’s favoured version of internalism goes under the label “the practicality requirement”, and says the following: Necessarily, if an agent has the opinion that it is right for her to \( \phi \) in circumstances \( C \), then she is either (i) motivated to \( \phi \) in \( C \), or (ii) practically irrational. (See The Moral Problem, p. 61. I have taken the liberty to change Smith’s own formulation so as to express more completely what he intends to say.) Practical irrationality as Smith conceives of it is a fairly complex notion, which includes such things as weakness of will and depression.

Actually, Smith defends two versions of internalism: “the practicality requirement on moral judgement” and “rationalism”. The latter claims that it is a conceptual truth that “If it is right for agents to \( \phi \) in circumstances \( C \), then there is a reason for those agents to \( \phi \) in \( C \)” (p. 62)
whereas most would accept that a good and strong-willed person is reliably motivated in accordance with her moral beliefs, fewer would accept that every rational person need be so motivated, although some have proposed elaborate arguments to that effect.\textsuperscript{1} In particular, externalists have thought that it is an open question whether we have good – rational – reasons to do what is morally right.\textsuperscript{2} (If Smith is right in his claim that beliefs about good reasons reliably cause motivation in the rational person, however, the externalist might say that the good person is one who believes that she has most reason to have non-derivative desires for moral values.)

The situation, then, is this. Changes in fundamental moral values seem to be reliably followed by changes in non-derivative desires in the good and strong-willed person. Internalism explains this connection with reference to the nature of moral opinions and the internal connection between moral grounds and moral motives. If the discussion above is right, externalists cannot explain reliability with reference to external desires; they should deny that the good person’s desires to act according to his moral opinions are created according to the following chart:

\begin{center}
\begin{tabular}{|c|c|c|}
\hline
External desires (1) & particular moral desire & \\
| & moral opinion (belief) + desire to do what is right & \\
| & moral reason (belief) (goodness) & \\
\hline
\end{tabular}
\end{center}

(Read: moral reasons give rise to moral opinions which trigger a desire to do what is right and thereby elicits a particular moral desire.) But externalists can refer to dispositions to form desires

\textsuperscript{1} If we consider various complicated arguments to the effect that rationality demands moral behaviour, such as David Gauthier’s \textit{Morals by Agreement}, we realize that the connection between rationality and morality is less than self-evident.

\textsuperscript{2} See David Brink’s \textit{Moral Realism and the Foundations of Ethics}, e.g. p. 50.
as a result of changes in fundamental values and argue that the good and strong-willed person is someone who – apart from having a number of non-derivative desires for such things as the well-being of others, honesty, justice, etcetera – has such a disposition to form new non-derivative desires for moral values, as represented below:

External desires (2)

<table>
<thead>
<tr>
<th>moral opinion (belief)</th>
<th>particular moral desire</th>
</tr>
</thead>
<tbody>
<tr>
<td>\</td>
<td>/</td>
</tr>
<tr>
<td>moral reason (belief) + disposition to form desires</td>
<td></td>
</tr>
<tr>
<td>(goodness)</td>
<td></td>
</tr>
</tbody>
</table>

(Read: when forming a moral opinion, the good person’s disposition to form desires is triggered by the same reasons that give rise to her moral opinion, thus resulting in a corresponding particular moral desire.) It is hard to see that internalists can deny the existence of such dispositions: on the internalist picture they are necessary for anyone to be able to have a dynamic set of moral opinions – such a set is a dynamic set of non-derivative desires. So, whereas Smith’s argument has added substantially to both internalist and externalist pictures of moral motivation, it has not been the sought for arbiter of the internalism–externalism issue.

3.6. Outline of an argument from simplicity

In this chapter we have seen how externalism is consonant with such things as our expectations that people will be motivated to act according to their moral opinions, our reluctance to think that it is morally wrong or right to perform actions that cannot be performed, and certain reasonable assumptions about the psychology of good people. Such facts have seemed to support internalism, but externalism can provide plausible alternative explanations by invoking desires to do the right thing or dispositions to form such desires.
But though the evidence put forth fails to show that internalism is correct it nevertheless makes the following plausible:

(1) Moral opinions are, at least frequently, accompanied by the corresponding optations. (Sections 3.1–4.)

(2) At least in the good and strong-willed person, the considerations that elicit moral opinions – her moral reasons – are those that elicit corresponding motivation. (Section 3.5.)

What internalism adds is, first, that all moral optations are like those of the good person, except that they might be weaker in some of us and that we sometimes stop our moral optations from influencing our actions (just as we sometimes stop other desires, such as cravings or impulses, to affect action), and, second, that moral opinions are nothing but such moral optations.

The problem is to substantiate this internalist generalization; the solution – to be developed in the following two chapters – is to show that alternative theories of moral opinions are arbitrarily complicated. In chapter 4 we will argue that:

(3) The externalist assumption that our moral psychology contains non-optational objective representations (“moral beliefs”, for short) adds nothing to our explanations of the phenomenology of the moral opinions of good people, or people that quite clearly have the relevant optations: the optations themselves account for the phenomena that moral beliefs could explain.

This does not, by itself, give us evidence that moral beliefs are missing when corresponding moral optations are present: (3) would do little to undermine externalism if (a) moral beliefs are needed to explain the phenomenology of moral opinions in cases where it is less clear that the relevant moral optations obtain and if (b) there is no reason to assume that cases where optations clearly do obtain are any different. What we will argue in chapter 5, however, is that:
(4) There are reasons to believe that the relevant optations obtain also in most of those unclear cases. Moreover, internalism can explain why we are inclined to think of certain cases as involving moral opinions even where it is clear that the relevant optations are lacking.

The upshot will be that there is no strong evidence for moral beliefs, but strong evidence for moral optations, and this supports the internalist claim that moral optations are what we think of and identify when we think of and identify moral opinions.¹

¹ On pp. 188–9 of his *Spreading the Word*, Simon Blackburn suggests a somewhat similar argument for “projectivism”, i.e. the claim that “evaluative properties are projections of our own sentiments” (p. 180).
4. Explaining signs of objectivism

Objectivism is an essential part of externalism, as defined in chapter 2. But although the truth of non-representationalism implies the falsehood of externalism, it does not logically imply that internalism is correct. However, if we were forced to give up the idea that moral opinions are states representing objective facts, we would have to face the question what they are instead; what they are if their function is not to keep track of moral reality. Here, the internalist can provide a general and fairly rich non-representational account, identifying moral opinions with moral optations: indeed, it is hard to see what a plausible non-representational general alternative would look like, and there seems to be none in the literature.¹ (And as we saw in the last chapter, externalists should acknowledge that good people at least have such optations: the internalist issue concerns whether moral opinions consist of such moral optations.) So, given that our goal is to find a general account of moral opinions, non-representationalism would provide a rather strong argument for internalism.

There is a serious problem with this line of reasoning, however: as far as I can see there are no convincing arguments for non-representationalism that do not assume internalism.²

¹ This does not imply that nothing less general could be said about what moral opinions are if we find objectivism and internalism incredible. We could point to and examine paradigmatic specimens of moral opinion without saying that what holds for them will hold for all states of mind correctly represented as moral opinions. See chapter 5 for a discussion of possible “deviant” cases of moral opinion.

² One of the problems with Allan Gibbard’s marvellous book Wise Choices, Apt Feelings is that his argument for a species of internalism seems to be founded on a problematic case against a form of representationalism, a case that builds on Moore’s ill-reputed “open question”-argument and a problematic argument from explanatory redundancy (“no need to assume moral facts”). (See pp. 3–23.) I agree with Nicholas Sturgeon in his criticism of these arguments. (See his notice of Gibbard’s book in Nous Vol. 29, Nr 3 1995, pp. 402–24.) The present argument is that there is no need to assume non-optational states of mind that represent moral facts: this is compatible with a need to assume moral facts.
What I believe, though, is that attention to arguments for objectivism could help us to decide the internalist issue. Of course, internalism and externalism alike are logically compatible with objectivism, and so it might seem that arguments for objectivism could do little for either theory. But moral psychology is not logic, and by the intricate line of argument suggested in section 3.6, internalism would be vindicated if it lets us explain phenomena that people have thought called for the positing of objective moral representations: externalism, but not internalism, would assume a kind of psychological state for which we had no explanatory need. Time has come to see what the relevant internalist explanations would look like.

4.1. The structure of the discussion
The task facing anyone wanting to adjudicate the issue between objectivists and their opponents is to identify properties in virtue of which an element in one’s mind is likely to represent some objective element in the world, and to decide whether (elements of) moral opinions have those properties. The problematic character of that task is made acute by the fact that optations are responsive to the world in various ways: in that respect they are similar to beliefs without therefore being representations in the sense needed for an externalist picture of moral motivation. Because a desire to perform a particular action depends on representations of that action, we should expect desires and beliefs to “behave” similarly although they have different proper functions, and that means that our instruments for distinguishing between them need considerable sharpness. Our survey of traditional arguments for objectivism will emphasize that point. What we will learn is that for every phenomenon supposedly indicating that moral opinions have an objectively representational function, an internalism uncommitted to objectivism has the means – many of which have been employed by authors such as Charles Stevenson, Richard Hare, Simon Blackburn and Allan Gibbard – to handle these arguments.
In the following sections, the discussion will more or less explicitly proceed through the following steps:

(1) A suggestion is put forth that a certain fact about our moral speech and thought is best explained by assuming that our moral opinions are objective representations.

(2) It is asked how this fact could be explained. Externalist explanations appealing to an objectively representational function of moral opinions are put forth and compared to internalist explanations appealing to a strategy-selecting function of moral opinions. What we will see is that internalism can provide explanations that are as plausible as their externalist counterparts.

Note that both externalist/objectivist and internalist explanations will have to invoke speculation about the functional organization of our psychological apparatus. In order to determine the relative merits of the two theories there is no way around speculating about what could reasonably be expected from our speech and thought given that our moral opinions are objective representations and moral optations respectively, and see which set of expectations match explananda better. Since supporting a theory with reference to its ability to explain certain data is a messy and empirical business, this means that results will quite possibly be controverted by further data. Moreover, there is no guarantee that moral opinions fall neatly on either side of the externalist–internalist distinction: at this stage of inquiry that is just a methodologically reasonable default assumption.

With these remarks in mind, we shall look at apparent signs of objectivism and show how an internalism uncommitted to objectivism – henceforth just “internalism” – can accommodate them.
4.2. Indicative surface

It is often observed that moral opinions are expressed or ascribed by means of sentences in the indicative mood, just as opinions which less controversially are about features of the world. Apparently without changing mood, we say that what he did was morally wrong, and that it was caused by greed; that attending the seminar was better, morally speaking, than going to the movies, and that it was less expensive. It might be claimed that the best explanation of this grammatical fact is that moral opinions represent features of the world.\(^1\)

One problem with taking this to support representationalism is that there are non-moral sentences sharing the indicative surface without obviously expressing beliefs rather than desires. Take Army orders for example. Or take what we think of as matters of taste: some people say that coffee tastes good, others that it does not; some say that reggae sounds better than blues, others would say the opposite. If I say, sincerely, that coffee tastes good, we might try to account for the state of mind that I express in a non-representationalist, a subjectivist, or an objectivist story:

- **Non-representationalism**: To be of the opinion that coffee tastes good is to like the taste of coffee.
- **Subjectivism**: To be of the opinion that coffee tastes good is to believe that one likes the taste of coffee; to represent oneself as liking the taste of coffee.
- **Objectivism**: To be of the opinion that coffee tastes good is to represent coffee as having a certain property, a property that is independent of whether one likes its taste or not.

It is far from obvious that the objectivist story is more plausible in this case, and that makes it far from obvious what the in-

\(^1\) That is how David O. Brink argues on pp. 25–5 in *Moral Realism and the Foundations of Ethics.*
dicative mood shows in the moral case. To get further we must ask why we would express our moral opinions by means of sentences in the indicative if we assume that objectivism and non-representationalism, respectively, are true. A plausible answer is that (1) we tend to follow inferential patterns in our moral reasoning (which is why we call it reasoning) and we tend to have conditional moral opinions (such as the opinion that if Charlie did something wrong, he ought to be told so) and (2) indicatives lend themselves better to inferences or conditional statements than, say, optatives or imperatives. (We do recognize something like inferential relations between imperatives too, but the imperative form is not well suited to handle inferences that concern acts that have already been performed, or the acts of people that cannot be addressed.) So, if inferential reasoning and conditional moral opinions are best understood from an objectivist perspective, that could support objectivism. In the sections to come, however, we shall see that internalism provides equally worthy explanations.

4.3. Simple inferences

Consider the following inference:

(A) All acts of hurting the innocent are wrong.
(B) Tom hurt an innocent person.
\textit{ergo} 
(C) Tom did something wrong.

It certainly seems that, from time to time, we make inferences of this kind. From the \textit{de dicto} premiss that all acts that exemplify a certain non-moral property are wrong and the premiss that a certain act exemplifies that property, we draw the conclusion that this particular act is wrong. And this practice of forming new opinions on the basis of old ones seems to be something that our moral reasoning has in common with reasoning involving paradigmatic cases of belief or representational states. Should we not suspect, then, that moral opinions are representational states too?
For a representationalist, the question of why we engage in simple inference is a question of why we form new moral beliefs about a particular act, starting from the belief that the act has some non-moral property and a belief that all acts having this non-moral property are wrong. And it is not hard to see what the point of forming such a belief would be, given that we often take a great practical interest in the moral properties of actions: given the truth of the general moral belief, the truth of the particular moral belief is guaranteed, and so is the wrongness of the particular action.

But the explanation is not harder to come by for the internalist: simple inferences are absolutely crucial for our general optations to have any import on action. If my optation for acts of hurting the innocent not to be performed is to have any purposeful effect on my actions, it must lead to the formation of optations for particular acts of hurting the innocent not to be performed. For that reason, the phenomenon of simple inference is evidentially neutral with respect to objectivism and non-representationalism.

4.4. Responsiveness to inconsistencies

If we are confident in our acceptance of premisses such as (A) and (B) of the inference of the last section, we often extend this confidence to a conclusion such as (C), and we think that anyone accepting such premisses but denying the conclusion is inconsistent in some way. In moral and non-moral matters alike, we charge ourselves and others with such inconsistencies and try to dissolve them by giving up one or more of the members of the inconsistent set of opinions. To be sure, we do not always correct inconsistencies, just as we do not always follow what, on reflection, most of us consider as sound inferential patterns, but we do so often enough for it to be a striking general feature of our moral thinking.\(^1\)

\(^1\) This is not to say that we are as faithful to consistency and standard inferential patterns in moral matters as we are elsewhere. My intuition is rather that
The usual way of explaining why we intuitively shun inconsistencies is to say that in any inconsistent set, at least one member must be false. Given that, and given that there is an interest in not believing what is false, our practice of avoiding inconsistencies would make perfect sense. Moreover, if we assumed that truth is correct mapping of or correspondence with reality, such an interest would seem perfectly natural: our interest in consistency could be seen as an interest in avoiding inconsistent maps of the world. It is hard to believe that beings moving around and surviving in a world like ours should not have such interests. So, if moral opinions are representations, there is a perfectly natural explanation of why we are interested in having consistent opinions. If they are strategy-selecting rather than representational, however, that explanation would not be available. Unless there is a good alternative, we have a nice argument for objectivism.

As many have pointed out, however, there is a non-representationalist explanation of our tendency to seek consistency that can adopt much of its representationalist counterpart. When we think of our beliefs as maps of various parts and aspects of the world, we get an idea of how beliefs can be inconsistent, and why that would matter: it would make our orientation in the world awkward. Now, instead of maps, we may consider such things as plans or blueprints. Blueprints do not represent what the world is like, but it is easy to see why we are more tolerant to inconsistencies in morals than in many other fields. An objectivist might want to accept this and try to explain it with reference to the great emotional and practical interest we take in moral matters, and the fact that it is hard to give up opinions around which one’s identity has been built. In any case, I will not use this intuition in any argument against objectivism.

1 In his “Outline of an Argument for Moral Realism”, Grazer Philosophische Studien Vol. 12/13 1981, pp. 215–25, Lars Bergström argues that there is no alternative explanation that would render our interest justified, and Torbjörn Tännsjö advances the same view on pp. 8–9 in Moral Realism, Savage, Md.: Rowman & Littlefield 1990. In the present context, however, explanation rather than justification is the primary issue (which is not to deny that the two sometimes go closely together).
would like to have consistent blueprints: if a set of blueprints is inconsistent, its members are not realizable together. (If we get several inconsistent blueprints of the house we are going to build, they cannot all be realized. Imagine, for example, that you are told to build the impossible house in one of M. C. Escher’s paintings.) But much as beliefs are maps of the world, optations to do this or that are blueprints of what to do or bring about. If our optations are inconsistent they cannot all be realized, which means that if moral opinions consist of optations – as non-representational internalism has it – moral opinions can be inconsistent in the same way. Thus interpreted, our interest in consistency and inference is explained by our interest in being guided by mental blueprints that are mutually realizable.¹

This account must be taken in the right way, though. In an argument against a non-representational account of our interest in inconsistency, G. F. Schueler argues that whereas it is a mistake to have inconsistent moral opinions, having conflicting attitudes is not:

Almost every time I walk past the cookie jar ... I experience such a conflict of attitudes. On the one hand, I approve of my eating a cookie since, from long experience, I know I will enjoy it. On the other hand, I also disapprove of my eating one since, from equally long experience, I know what it does to my waistline. So here is a clash of attitudes, but where is the “mistake” ... that is supposed to be involved?²

¹ See especially Simon Blackburn’s Spreading the Word, chs. 6 and 7 and “Attitudes and Contents”, pp. 190–191 in his Essays on Quasi-Realism, but also Allan Gibbard’s Wise Choices, Apt Feelings, pp. 74–5, 284–91 and Folke Tersman’s “Non-Cognitivism and Inconsistency” in Southern Journal of Philosophy, Vol. 33, 1995, pp. 361–371. In his The Logic of Commands, London: Routledge & Kegan Paul 1966, Nicholas Rescher argues that the logic of commands should be approached by means of ordinary assertoric logic and the notion of command termination, i.e. the realization of the action commanded. (And of course the function of commands, like that of desires, is to cause action.)

Agreed, although the clash may be unfortunate, there is not much by way of mistakes involved in it. But neither do we think that there are mistakes involved in similar moral clashes: we sometimes say that an act would be morally wrong in one respect (being a breach of promise, say) but right in another (saving an innocent person from great social embarrassment, say) and see nothing mistaken in holding these “conflicting” moral opinions. It is when we have the opinion that the act is morally wrong all (relevant) things considered that we think of it as a mistake (or perhaps an impossibility) to be of the opinion that, all (relevant) things considered, not performing the act is morally wrong, too. (Whether, and in what senses, there might be “real” moral dilemmas – situations in which one cannot avoid making something wrong – is of course a difficult matter that cuts across the border between normative ethics and meta-ethics. The point here is just that certain moral clashes appear to be as innocent as certain clashes of desires.1)

One way to see how non-representationalism can explain our limited tendency to shun inconsistency is to ask when it would make sense to form the kind of optations that the non-representationalist identifies with such all-things-considered opinions. The most obvious case is when one is forming an intention to act, and when the resulting moral optation causes (or perhaps is) an intention to act:2 on such occasions it is quite

2 Intentions, like desires, seem to be somehow connected to or have the function of eliciting action – in addition to being formed when one decides upon some kind of action and so forms the belief that one will perform that act. Several analyses of intentions seem to be possible here. We might, for example, identify the intention with (1) a kind of belief that one will perform a certain act, a belief that is caused in non-deviant, functionally normal, ways by a desire to perform that act, or (2) with the desire that causes such a belief, or (3) with the conjunction of the belief and the desire, or (4) with a state that is both a belief of the relevant kind and a desire. (See Ruth Garrett Millikan’s “Pushmi-pullyu Representations” for the latter view.) Various amendments of the above are also possible and perhaps called for. Gilbert Harman, for example, thinks that an inten-
easy to see why inconsistencies would lead to trouble and why we would have a tendency to remove inconsistencies when becoming aware of them.\(^1\)

Of course, we have moral opinions about a number of other cases too, and it might be objected that our interest in avoiding inconsistent moral optations cannot explain why we care about consistency even between opinions concerning the actions performed by others or long before our time, or merely imaginary actions.\(^2\) In answering this objection, the non-representationalist can follow two complementary strategies, asking us to hold in mind that many people do not care a great deal about consistency in their moral opinions about such cases:

(1) He can point out that the representationalist has the same problem. We care about the consistency of beliefs that concern very remote and practically uninteresting parts of world: why care whether maps of those parts are consistent? It might be retorted that it is the function of beliefs to be true, or correspond to the facts – that is what they are supposed to do, regardless of what parts of the world they concern.\(^3\) But it is not clear what that would show since the non-representationalist

\(^1\) See pp. 189-91 of Simon Blackburn’s “Attitudes and Contents” in his Essays on Quasi-Realism for a similar answer to Schueler.

\(^2\) In “Outline of an Argument for Moral Realism”, Lars Bergström seems to think that our interest in consistency exceeds that in avoiding inconsistent blueprints. Walter Sinnott-Armstrong takes a slightly different approach in his criticism of non-representationalist accounts of our intuitions concerning inconsistency, claiming that not all cases of inconsistency are the subjects of pragmatic pressures, despite its still being something wrong with them. According to Sinnott-Armstrong, this leaves the validity of arguments containing normative judgements without full explanation (“Some Problems for Gibbard’s Norm-Expressivism”, pp. 301–2).

\(^3\) If we like to think of functions as something states or things do, rather than what they are, we can put it like this: in order for beliefs to fulfil their function – to allow us to act so as to achieve our goals – in a non-coincidental manner, they must correspond to the facts, or be true.
could say, similarly, that it is the function of optations to lead
to the realization of the desired kind of state.

(2) The non-representationalist can give reasons to expect our concern for consistent blueprints to be quite extensive. One reason to expect an interest in consistency that transcends one’s own present action is that we are interested in the doings of others. If our advice, commands and requests are to be followed they had better be consistent, since a set of inconsistent commands cannot be carried out. Moreover, it would be difficult to cooperate with a person with inconsistent moral optations in so far as her moral optations lead to action.\(^1\) (To some extent, the demand for consistency holds equally for inconsistency over time, since a person who frequently changes his advice or his commands looses authority and is incapable of forming and realizing complex projects.) Finally, some of our judgements about imagined or spatio-temporally remote cases make sense if seen as rehearsals for our own decision making and commands, preparing us for actual cases just as role-playing can prepare lawyers, business-people, and military personnel for their future practice.\(^2\)

Now, given that an ability and tendency to increase optational consistency would make sense for the homely cases, it is likely that the mechanism involved will apply equally well when cases without immediate or even the least likely pragmatic consequences are concerned. Most of our practical thinking does concern behaviour that is pragmatically interesting in the ways sketched above, and it is hard to see how the costs involved in checking and revising opinions for consistency in a small

---


\(^2\) Here I agree with Allan Gibbard’s suggestion in *Wise Choices, Apt Feelings*, pp. 74–5.
amount of “useless” cases would outweigh the costs of discrimi-
nating (in advance) between cases of different pragmatic rele-
vance.¹

It seems, then, that the non-representationalist has the
means to account for our tendency to give up inconsistent (not
mutually realizable) sets of moral opinions, at least in simple
cases. But there are decisive problems with a general identifica-
tion of inconsistency among moral opinions with lacking mutual
realizability. One such problem appears when we consider cer-
tain conditional or disjunctive opinions; another when we look
at permissive moral opinions, such as the opinion that a certain
act is neither morally wrong nor what ought to be done. In the
following two sections, however, we shall see how non-repren-
sentationalism can provide solutions.

4.5. Unasserted contexts: disjunctions and conditionals

Suppose that someone utters the following sentences:

¹ This is also what should be said about beliefs. It is true that our concern for
consistency among our beliefs has a wider scope than the set of pragmatically
interesting beliefs – those on which we are likely to act – but it is only to be
expected that evolution and culture should give us a capacity and tendency to
check for and eliminate inconsistencies that go beyond that, since pragmati-
cally uninteresting actual beliefs are relatively few and hard to pick out in
advance. (The last point becomes especially clear if we consider that one way to
act on a belief is to inform someone else of it, as we do when we describe events
that are geographically or temporally remote. The mere fact that someone else
is interested in some obscure fact can make my belief about it pragmatically
relevant.)

In his “Ethical Consistency” and “Consistency and Realism” (both in Prob-
lems of the Self, London: Cambridge U. P. 1973), Bernard Williams argues that
what we have called objectivism gives moral inconsistency an “ultimate kind of
significance” whereas non-representationalism would give it limited pragmatic
significance. Williams is quite right in his claim that inconsistency among
desires (or imperatives) is different from inconsistency among beliefs, but I
suggest that his talk about “ultimate significance” expresses a philosopher’s
penchant for truth rather than any insight into the matters at hand. Someone
with a more pragmatic bent might see mutual realizability as ultimately signifi-
cant, taking the pursuit of truth and consistency among beliefs as a means to
that goal.
(D) Either lying isn’t wrong or making one’s little brother lie is wrong.

(C) If lying is wrong then making one’s little brother lie is wrong.

It seems that these sentences express neither the utterer’s opinion that some act or kind of act is wrong, nor any desire that an act or kind of act be refrained from. Still, we do form moral opinions expressed by disjunctive sentences like (D) and conditionals like (C), opinions that figure in inferences and are intuitively taken to be consistent or inconsistent with other opinions. According to a line of argument that was inspired by Frege but found its way into meta-ethics through P. T. Geach’s 1965 paper “Assertion”\(^1\), this poses a decisive problem for non-representationalist theories of moral opinion. For simplicity, and since (D) and (C) often mean the same, we shall focus our discussion on (C).

In considering this argument, it is important to understand just what the difficulty here is. It is not, as far as I can see, to give a general (normative) theory of how (C) partakes in valid inference and how it stands in normatively significant relations of consistency and inconsistency to other sentences in virtue of its particular form. According to our intuitions about such question, (C) presumably works just as any conditional sentence with clauses in the indicative mood. The occurrence of the predicate “wrong” – which turns (C) into a sentence conventionally expressing a moral opinion of sorts – does not seem to change the patterns along which the resulting conditional works in valid inference and is consistent or inconsistent with other sentences.\(^2\) Consider the following sentences:


(C*) If lying makes me nervous then making my little
brother lie makes me nervous.

(L*) Lying makes me nervous.

(M*) Making my little brother lie does not make me ner-
vous.

Intuitively, they form an inconsistent set, but so does (C)
together with (L) and (M):

(C) If lying is wrong then making one’s little brother lie is
wrong.

(L) Lying is wrong.

(M) Making one's little brother lie is not wrong.

This is not the place to decide whether our normative intuitions
are correct or well founded (although arguments for and against
objectivism would seem to have quite some relevance for that
issue). The present problem, rather, is to explain that apart
from “categorical” opinions such as the opinion that making
one’s little brother lie is wrong or that telling lies is wrong, we
have what we shall call “conditional” moral opinions: the kind
of opinions expressed by conditionals where a moral sentence
fills the “if”-clause, such as (C). In other words, what needs an
explanation is that we have opinions that lead us to

(1) form new moral opinions as a result of (being aware
of) having other moral opinions, as when having the opin-
ion that lying is wrong leads us to form the opinion that
it is wrong to make one’s little brother lie, and

1 Conditionals with a sentence expressing a representational opinion in the
“if”-clause pose much less of a problem for the non-representationalist, since
they can be seen as straightforwardly expressing an optation. For example, the
opinion that it is wrong to lie if it is done to save oneself from a minor discom-
fort can be seen as a desire that lies done to save oneself from a minor discom-
fort not be told.
(2) give up one moral opinion as a result of (being aware of) having other moral opinions, as when having the opinion that lying is wrong leads us to give up the opinion that it is not wrong to make one’s little brother lie.¹

Before we look at what explanations the competing theories can offer, we shall note that conditionals can be of very different kinds, expressing connections of different modalities between the antecedent and the consequence. That means the we should expect slightly different explanations of different conditional opinions. For the sake of simplicity, however, we confine ourselves to a discussion of two examples, betting that further investigation would not tip the evidential scales. Start with the tautological opinion that:

\[(C') \text{ If an act is wrong then it is not the case that it is not wrong.}\]

To be of the opinion that \((C')\) is to be disposed to give up and refuse to adopt inconsistent categorical opinions concerning the wrongness of particular acts, at least when not suffering from irrationality, and when such inconsistencies are manifest. If the opinion that \((C')\) can be said to have a function, it is to avoid and get rid of inconsistencies. The question, then, is how to explain that someone might be in a state of mind the function of which is to get rid of inconsistencies. Now, as we saw in the discussion above there is a ready representational account of our interest in consistency, but we also saw that the non-representationalist could provide a plausible alternative. So, if the rep-

¹ Here I agree with Simon Blackburn’s “Gibbard on Normative Logic” in Philosophy and Phenomenological Research Vol. 52, No. 4 1992, pp. 947–52. By contrast, the primary problem posed by Geach and those inspired by his argument is to account for the validity of inferences in which disjunctions or conditionals figure as premises without recourse to truth-functional logic. What I consider as data here, however, is not the validity of certain inferences, but our tendency to give up or form certain moral opinions upon accepting others. The task is to explain this tendency.
resentationalist theory has any advantage here, that must be because it provides better explanations of more substantial conditional opinions, such as the opinion that:

\[ (C'') \text{ If acts of a certain kind are wrong then making someone perform an act of that kind is wrong.} \]

Presumably, a person can be said to be of the opinion that \((C'')\) if she is disposed to

1. form a moral opinion to the effect that making someone perform an act of a certain kind is wrong, and
2. give up opinions to the effect that making someone perform an act of a certain kind is not wrong, when

\[ (2) \text{ having the opinion that acts of that kind are wrong.} \]

The question to ask is whether being in such a state is equally intelligible on non-representationalist as on representationalist theories.

On representationalism, there would be a point in having the opinion that \((C'')\) if the structure of moral reality was such that if acts of a certain kind have a certain property – moral wrongness – then making someone perform an act of that kind has that property too. Under such circumstances, the opinion that \((C'')\) would be a disposition to adjust our moral beliefs so as to correspond to that reality – given the truth of our beliefs about what acts are wrong. Moreover, since the conditional opinion can fulfil this function of adjusting our categorical moral beliefs to moral reality in a non-contrived or purposeful way only if moral reality has the relevant structure, we could say that the conditional belief represents that structure or that it is supposed to correspond to such a structure, or that it is true if that structure obtains. So, since it is to be expected on an objectivist view that we should have opinions mapping interesting structures of moral reality, that view fits well with our practice of
making conditional moral statements and having conditional moral opinions.¹

The non-representationalist cannot assume that (C"') is supposed to correspond to, or map, some regularity of a represented world, of course. But there are other ways of explaining why someone might be in the state of mind such as that expressed by (C"'). Assume that a person’s moral optations exclude certain acts (those which she regards as morally wrong) because these acts are perceived to have certain consequences or result from a kind of motivation that she dislikes. Now, making someone else perform acts having such consequences or resulting from such motivation is to cause such disliked consequences or motivation. It is therefore perfectly intelligible that when (she is aware of) having the optation that acts of a certain kind not be performed, she

(1) forms a moral optation that acts of making someone perform an act of a that kind are not performed, and

(2) gives up moral optations that acts of making someone perform an act of that kind be performed.

And that is, in non-representationalist light, the disposition that constitutes having the opinion that (C"')

Of course, that a state of mind is intelligible under certain conditions does not mean that everyone having a moral opinion under those conditions is in that state. A consequentialist, for example, would not accept (C"') since causing someone to perform an act with sub-optimal consequences might itself have optimal overall consequences. A complete explanation of why someone has the opinion that (C"') would have to invoke some difference between that person and the consequentialist. Moreover, although the above explanation of why someone might

¹ The question why anyone would adopt (C") in particular is a question of why anyone would come to believe that moral reality has the structure represented by (C"). But trying to answer that question would take us too far since any answer would refer to other beliefs of that particular individual.
have the opinion that (C") illustrates how a conditional opinion is the effect of a certain structure in a person's categorical opinions, the explanation of why a particular person's moral blueprints have a structure that can be expressed by a certain conditional moral sentence will vary from case to case. The point is that there is no obvious reason to assume that such explanations are harder to come by for the non-representationalist than for the representationalist.

The upshot is that non-representationalism provides a worthy alternative to representationalism when it comes to accounting for conditional (or disjunctive) moral opinions. At least that has seemed to be the case with the examples that we have looked at, and we have no reason to expect other examples to tip the scales in favour of either theory.

4.6. Permissive opinions and consistency

Intuitively, the opinion that an act is morally optional or permitted is inconsistent with the opinion that it is morally wrong. But how should we understand such inconsistency, or our interest in it?

If objectivism is true, that seems to provide no special problem. Moral wrongness is believed to be such that acts cannot exemplify it when they exemplify the property of being morally obligatory; the relation between wrongness and the property of being morally permitted is believed to be just the same. For a non-representationalist, however, the matter is more difficult. Think about what can happen on a non-representationalist view when a moral opinion about an act is formed. A person considers what he takes to be facts about the act, and this triggers certain strategy-selecting states, resulting (perhaps under the influence of second-order optations) in either (1) the moral optation that the act be performed, or (2) the moral optation that the act not be performed, or (3) neither of the above — he is morally neutral. (Let us assume that the case is not dilemmatic.) The first result is the moral opinion that, morally, the act must, or ought to be, performed; the second is the moral
opinion that it is morally wrong to perform the act. The third result, however, seems to be no new state of mind at all, and thus nothing with which to identify a permissive moral opinion. That, then, is the problem. But as we will see presently, it is reasonable to expect even this negative result to leave a positive trace since it can be of considerable practical interest.

In order to highlight features of the practical reasoning of an individual it is often useful to think of the decision-making of a collective such as a nation. One thing that is striking about collective decisions and which holds equally at the level of the individual is that decisions take time: this is explained (at least partly) by the fact that the information available to the deciding agent is spread out in space. Whether the collective decision-making procedure ends in one individual’s deciding for the group or by some kind of voting, information has to be brought to the attention of the chairman or the meeting in order to have effect on the outcome. Correspondingly, an individual agent is not immediately aware of all that she knows when trying to decide, but has to search in her memory for relevant facts. Because of this, and because only a limited amount of information can be kept in working memory, it is important for the agent not only to remember goals represented by decisions long enough to achieve them (for which she needs to save positive and negative results of her moral deliberation) but also to remember results that allow her to forget certain aspects of the alternatives and concentrate on others. And keeping her from further considerations of facts relevant to moral optations seems to be exactly what permissive moral opinions allow her to do.

While this shows that we can expect negative results of practical deliberation to change a person’s psychological states, it remains to be explained why the resulting states would be inferentially related to moral optations in the way that permissive moral opinions are related to opinions of moral wrongness or obligation. The explanation is quite straightforward, however. Being in a state the function of which is to keep one from considering moral aspects of a kind of actions (having the opin-
ion that such actions are permissible) while at the same time being in a state the function of which is to lead one to act from exactly such consideration is likely to cause trouble. First, it can make one’s deliberation unstable and thus more time-consuming, since having the permissive moral opinion would push one towards one type of deliberation while having the obligating moral optation would push one towards another type. Second, unstable deliberation makes long-term planning less worth the while since the resulting action of such deliberation will be hard to predict. Third, cooperating with an agent who lacks stable patterns of decision and action is difficult, which makes for social pressures not to have both a permissive and an obligating opinion concerning the same kind of actions. Holding these things in mind, permissive moral opinions are no mystery for non-representationalists.

4.7. Mistakes and disagreement

When we change our minds on moral issues, we typically come to think that our former view was mistaken, just as we do when we change our minds on paradigmatically factual issues. If a person formerly had the opinion that some acts of euthanasia were morally right but changed her mind, now being of the opinion that no cases of euthanasia are morally right, she would probably say that her former view was mistaken. And the same holds about cases where our present moral opinion is inconsistent with the moral opinions of others, rather than with the moral opinions of our former selves: we tend to think that their views are mistaken. On top of this, we tend to think that some of our own moral opinions might be mistaken.

On the objectivist story, such ascriptions of mistakes seem to be perfectly understandable, since at least one of a pair of inconsistent representations must be false, and therefore mistaken in a straightforward sense, and since moral facts can be

1 Of course, this is not so obvious with subjectivist representationalism, since it takes the truth of my moral opinions to depend on facts about my idiosyncratic
epistemically intractable.\footnote{See section 4.1.} The non-representationalist picture cannot be that simple, though, as it seems perfectly possible that a desire of mine and a desire of yours are inconsistent (not mutually realizable) without either of them being mistaken in any interesting sense: think of the U. S. Open finalists who both desire victory.

Obviously, we cannot take as a datum that inconsistency among moral opinions implies mistake in the sense of falsehood. And although many have the intuition that moral inconsistencies imply mistakes in some way, others – notably non-representationalists sympathizing with Humean anti-rationalism – do not, and my experience with people innocent of philosophy is that a considerable number can be quite easily goaded or seduced both into accepting and into denying the implication between inconsistency and mistake. What we have as data, though, is (1) a relatively wide-spread (but certainly not all-pervading) practice of trying to dissolve disagreement in moral matters by means of argument, (2) quite uniform intuitions that one’s present moral opinions represent an improvement upon earlier ones, (3) quite widespread intuitions that at least some of one’s own present moral opinions might be mistaken, and (4) quite uniform intuitions that moral opinions that are inconsistent with one’s own are mistaken. In his argument, the representationalist might want to say that we behave and intuit “as if” moral opinions were representations, and that the best explanation of this is that moral opinions are representations. Let us see whether he would be right in saying that.

(1) **Dissolving disagreement by argument** Representationalist explanations of our efforts to dissolve disagreement by means of argument seem to be ready at hand. Given what seems reasonable under objectivism, namely that we want ourselves and

\[\text{optations whereas the truth of your moral opinions depends on facts about yours.}\]
others to know and do what is right and avoid what is wrong, it is only to be expected that we try to decide which of two inconsistent moral opinions is correct by means of argument, and try to convince others when we think that we know the truth. The non-representationalist, on the other hand, can invoke the fact that we (a) care about what moral optations other people have — especially when these optations concern actions that are important to the well-being of ourselves and our near and dear — and (b) believe that we can change their moral optations by means of argument. Hence it is only to be expected that we will try to do so. Moreover, the non-representationalist can explain why our optations can be changed by argument. Although other parts of that explanation will emerge when we look more closely at the rival explanations of (2) and (4), the most important reason is that moral optations are elicited by beliefs about the act in question, and such beliefs can be mistaken and more or less complete. There is no controversy about the possibility of changing such beliefs, and changing optations with them.

(2) Intuitions of improvement Why do we tend to think that, after having changed our minds on some moral issue, our present moral opinions improve upon our earlier ones? One objectivist answer is that if we now believe that an act is wrong we are likely to realize that this implies that what we believed before (namely that the act was not wrong) was false, and thus mistaken. The problem with this explanation is that it invokes the belief that one’s former opinion was false, rather than the fact that it was false (and hence that it was representational), but not everyone believes that moral opinions can be true or false, and the question asked in this chapter is why we should

---

1 See section 4.2.
2 Stevenson’s Ethics and Language is the classical work on non-representationalism and issues of disagreement and argument. See especially chs 1, 5, 6 and 11. A modern writer in the same tradition is Simon Blackburn, for example in Spreading the Word, ch. 6, § 2.
believe they can. A better explanation goes like this: assuming that our representations are more likely to lead us right than are their negations and assuming that we have a tendency not to knowingly adopt inconsistent sets of beliefs, it is only to be expected that we have a general tendency not to adopt, once again and without new information, beliefs that we have given up for our present beliefs. According to the objectivist, this reluctance to adopt a former belief is what causes our intuition or feeling that our former moral opinion is worse than our present.

The problem with this objectivist explanation is that the non-representationalist can offer something very similar. There is an evolutionary rationale for having the strategy-selecting states we have rather than their opposites: the strategy-selecting systems that created them have been selected for their tendency to go into states that led to the propagation of their genes. Moreover, and as we saw in section 4.4, non-representationalism should lead us to expect a tendency not to knowingly adopt inconsistent sets of optations. Hence, it is only to be expected that we have a general tendency not to adopt, once again and without new information, optations that have been given up for our present optations. It is this reluctance that the non-representationalist thinks of as causing our feeling that our former moral opinions are somehow worse than our present ones.

In particular cases we might not only be reluctant to adopt our former view, but we might also be of the opinion that there are better reasons for having our present moral opinion than for having our former one. Suppose for example that your disagreement with your former moral opinion is based on improved knowledge of non-moral facts and improved reasoning. As we said in section 3.1, it is only to be expected that moral opinions can be more or less well founded if objectivism is true, and it would seem equally reasonable that we have ideas about how well founded our moral opinions are, ideas that partly concern the truth and completeness of our moral grounds and partly the methods that we use in assessing these reasons. The question
is whether non-representationalism can explain that we think of our moral optations as more or less well founded.

To provide such an explanation, we shall look at various ways in which we might come to think that moral optations are mistaken, in the sense that we would not equally well have the one as the other. The question, then, is whether we should expect ourselves to have second-order optations about what moral optations to have. With two kinds of second-order optations this seems especially obvious. (A) We do not want to have optations based on a belief that we think is false: if I desired not to perform an act because I believed that it would belittle someone, but now realize that the act would not belittle that person, I now desire not to have my former desire. (B) The belief that a desire is based on more information than a preceding and conflicting desire leads us to desire to have the later rather than the former one. If I desired to perform an act but learned that it would belittle an innocent person and therefore no longer desire to perform it, I now desire not to have my former desire.

So, second-order optations and corresponding intuitions about mistakes could be explained by our interest in having desires based on all relevant and no false information. It seems clear, though, that this falls short of letting the non-representationalist mimic representationalist explanations of our intuitions. Consider that our moral opinions are the effects of the weight we give to various facts about actions (that they cause suffering, infringe on peoples’ autonomy, belittle someone, and so on). The non-representationalist cannot say, with the objectivist, that giving a certain weight to one fact about an act and another weight to another fact in assessing its moral status is to make a representational judgement about how these facts contribute to the moral quality of the act. But if weighing is not a representational judgement, it seems possible that two persons can weigh the same facts about an act differently without either having based their weighing on false beliefs. At the very least we would need a good argument to deny that possibility, and I have never heard of any remotely plausible candidate.
The non-representationalist might insist that even if the explanation above fails to give the same scope for mistakes as its representationalist alternative, it gives the right scope: when there has been no improved knowledge about the act in question, we do not think that our former opinion was mistaken or that the new opinion represents an improvement upon the earlier one. But I doubt that this response goes well with our intuitions: we probably feel that our moral opinions have improved even when the change has no clear ground in improved factual knowledge about the act in question.

What the non-representationalist should do, however, is to insist that there are other ways in which we might come to feel that our moral optations / opinions have improved. For example, since we strive for consistency, the realization that our present set of optations is more consistent than our former set would be taken as an argument in favour of our present set. (To see how much this covers, it is important to remember that we often have conditional moral opinions.) Consciously or not, we might also feel more confident with our optations if they (and resulting actions, including linguistic expression of these optations) promise to go better with the expectations and attitudes of others, allowing us to feel at home in our community. That is probably especially – though not exclusively – true when young children are concerned. (It is also, I think, an explanation of why we sometimes feel that mistakes are very real in matters of humour and style, even though it is hard to believe that those are matters of objective fact.)

Reference to social pressures can also be invoked by the non-representationalist in explanations of why feelings of improvement are more strongly invited by disagreement and change in moral optations than in others – something which both non-representationalists and objectivists need to say something about. Part of the explanation will be that moral optations concern actions, and that actions can be changed if optations (desires) are changed. If Charlie desires rain tomorrow, but Ada does not, there is typically nothing important for Charlie to gain by
changing Ada’s desire: it will not make it rain. If Charlie desires for Ada not to have an abortion but Ada desires to have one, however, Charlie might fulfil his desire by changing Ada’s desire, so that she gives birth to their child. But it is also important that our reasons for moral opinions often concern matters that we care much about, such as the distribution of goods and the respect and welfare of others. All this means that the social pressure concerning moral optations will tend to be stronger than that concerning most other optations.

So, we can explain why we feel that our new moral opinions represent an improvement upon our former ones with reference to (a) the possibility of improved knowledge, (b) improved consistency among one’s moral optations, and (c) improved consistency with non-moral desires to fit in or to otherwise feel good about ourselves. Let us therefore turn to the last two explananda.

(3) Intuitions of fallibility Since the above stories (whether objectivist or non-representationalist) explain how and why we find some moral opinions mistaken, it is not hard to understand why we expect to find some of the moral opinions that we now take to be correct to be mistaken: all it takes is simple induction.¹

(4) Intuitions concerning the optations of others When I find that someone has a moral opinion inconsistent with my own, my first reaction tends to be that he is mistaken (depending on who this other person is, and what the issue is). Others seem to have similar tendencies. But how should these tendencies be explained? For an objectivist, the answer appears to be fairly straightforward: I see the other as mistaken in so far as I continue to have my own moral opinion, since the truth of my own

¹ For an early attempt to show how non-representationalism is compatible with fallibility, see Charles Stevenson’s “Ethical Fallibility” in Ethics and Society, ed., Richard T. De George, London: Macmillan 1968.
opinion implies the falsehood of his. But as with the intraper-
sonal ascription of mistakes discussed above, it cannot be
taken as a datum that any realization that you and I disagree
in our moral opinions leads us to think that the other is mis-
taken in the sense of having a false belief. Relativists and sub-
jectivists of various sorts have denied this at great length, of
course, but whether they are right or not most of us certainly
tend to think (however vaguely or out of place) that there are
better reasons to have our own moral opinion. In some cases,
this can be a rather empty reason: I might think that it is a
great mistake for any rational being not to have the opinion
that all acts of deliberately killing the innocent are wrong just
because it is that opinion, or (proudly) because it is my opinion.
In other cases, the reason for thinking that someone is mis-
taken in having a moral opinion can be that the opinion is in
conflict with other moral opinions (which may or may not be
considered as being held for good reasons) or that it is based on
false assumptions or ignorance. The question, then, is why we
should expect a person to have ideas about when others should
or should not have certain moral opinions: the interpersonal ver-
sion of the question that was answered by both objectivism and
non-representationalism under (2).

The objectivist can point to the fact that we do have such
ideas concerning paradigmatic beliefs about objective features of
the world. We frequently think that others are mistaken about
such mundane facts as where they put their sweater, what the
quickest mode of transportation between two cities is, or what
team will win the Super Bowl; and we have various ideas about
under what circumstances one is better placed to have correct
beliefs about these things, as well as ideas about what people
to trust in various matters. The harder question seems to be
whether non-representationalism can account for these ideas.
Now, since non-representationalism takes our thoughts about
what moral opinions to have to be second-order optations about
moral optations, the question is whether it makes sense that
we have second-order optations about the moral optations of
others. And it seems most natural that we do: such optations make us issue commands, blame and praise, and thus affect the acts and attitudes of others, and so affect our own lives. Among the things that we can be expected to care about regarding the optations of others — except that others do not desire to hurt us or our projects in various ways — are that their optations be consistent and based on good information. The reason for this — which was mentioned in section 4.4 — is that an agent who is inconsistent in her optations or fails to apply her general optations to particular cases due to serious misapprehension of the facts will be a bad partner for the kind of cooperation that human societies with all their boons are built upon.

With that said, I think we have seen that non-representationalism can produce satisfactory explanations of our intuitions about mistakes. What we have not done is to supply explanations of why particular agents come to have particular ideas about mistakes; but that is equally missing in the objectivist account above, and not of obvious interest here. We shall therefore turn to other putative signs of objectivism.

### 4.8. Uncertainty

Often, we are uncertain about whether an act is morally right or not, and we sometimes want to know which. If moral opinions are representations, this seems easy to explain: it is hardly surprising that we sometimes want to learn more about the world. Of course, we are not interested in every part and aspect of the world, but as we noted in chapter 3, it seems very plausible that if our moral opinions represent moral facts, then such facts are interesting from the point of view of practice. But what can the non-representationalist say?

An intuitive answer is this: uncertainty with respect to the moral status of an act is lack of determinate moral optations concerning the performance of the act. But one might think — with Lars Bergström — that this cannot reasonably account for
the fact that we want our uncertainty dissolved: that we want answers.\(^1\) Why should we want to have either the optation that the act be done or the optation that it not be done, or the state of mind constituting a permissive moral opinion? That is a good question, but we will see that there are equally good non-representationalist answers.

To begin with, most of us do not have a strong desire to know the answer to any arbitrarily picked out moral question. But in the kind of case where we are most interested in the matter – when we are in the position to perform the action in question, or to give advice, or when we try to understand someone else who is in such a position – the question of what opinion to have is a question of which action to perform, or what to (morally) want someone else to perform, and those are intelligible questions. In such cases, it is only to be expected that we want to have optations that satisfy certain conditions, such as being consistent with other desires and being grounded in true beliefs. (Admittedly, some people – professional philosophers, for example – take an interest in answers to moral questions that go beyond the homely, but such idiosyncrasies will have to be explained with reference to facts about such people, not primarily with reference to general facts about moral opinions.)

Another problem suggested by Bergström is posed by the fact that whereas optational strength seems to have but one dimension – the optation’s functional tendency to lead to action – our moral opinions can vary in two dimensions: in doxastic and moral strength.\(^2\) The doxastic strength of an opinion is the degree of confidence with which it is held, whereas the moral strength of a moral opinion is the degree of wrongness or obligation it attributes to the act in question – whether it is very wrong or merely somewhat wrong, or whether the moral obligation is strong or weak. Now, the non-representationalist cannot understand doxastic strength as the confidence with which the


\(^2\) *Op. cit.*, pp. 35–6
opinion is taken to correspond to moral reality, of course, but it can plausibly be seen as the person’s confidence that the opinion is not mistaken, and it was shown in section 4.7 above how our talk about mistakes can be given a non-representationalist reading. If there is a problem here, it concerns the non-representationalist’s account for our intuitions about moral strength.

Here our intuitions about the moral strength of an opinion can be seen as intuitions about the optational strength it would have had if the doxastic weakness was factored out, that is, if the person had been completely confident in his opinion. In a way, that account might not be satisfactory since we seem to lack precise or determinate intuitions about optational strength. But since our intuitions about moral strength seem to be afflicted with the same looseness, that is how it should be. And so it seems that the non-representationalist is able to account for our uncertainties, too.

4.9. Explanations

Sometimes we might want to say for example that a person’s fall from office was caused by morally bad conduct. Moreover, as Nicholas Sturgeon has pointed out, if we learn that our moral opinions are caused or sustained by reasons that have nothing to do with the actual rightness or wrongness of the actions they concern, we seem to be left without grounds for holding them:¹ it seems that the possibility of explanations in terms of moral rightness and wrongness is essential to our confidence in our moral views. But if reference to moral rightness or wrongness is to provide any real explanation or prediction of events or of our moral opinions – as we surely intend it to – talk about morally wrong actions must pick out something in the world that can cause things to happen: cause people to fall from office and us to have moral opinions. That is exactly what opinions representing objective features of the world would do, but how can

non-representational opinions pick out the relata of causal relations?

A very plausible answer departs from the fact that when we are of the opinion that some act is morally wrong, this is because we believe that the act has certain properties other than its moral wrongness. Even if there are no moral facts, the fact that the act has those non-moral properties can have causal effects. Suppose for example, that the chairman of a company steals something from the owner of the company and is fired as a consequence: the owner does not want a chairman that cannot be trusted. If we are of the opinion that the chairman’s act was wrong on the ground that it was a theft and that it was done for selfish reasons, we might want to say that morally wrong conduct, or the moral wrongness of his conduct, caused the chairman’s downfall. The fact that it was a theft done for selfish reasons caused the owner to distrust the chairman and so to fire him.

It must be stressed that there is nothing ad hoc about this answer. The above sketch of what leads us to make explanatory judgements in terms of moral properties is compatible with both non-representational and objectivist theories. Moreover, it seems quite plausible on an objectivist view, since it is hard to see that the instantiation of a moral property should have any causal effects apart from those resulting from the instantiation of the non-moral properties that constitute the moral property or make it the case that the act has the moral property. For example: even if it is a fact that the chairman’s act of theft is morally wrong, it is hard to see that the moral wrongness of the act should have any causal effects apart from those of the facts (perhaps taken together) that the act was a theft, that it was done for selfish reasons, and whatever else makes the act wrong or constitutes its moral wrongness. The fact that these properties make it wrong or constitute its moral wrongness does

1 See pp. 205–6 of Simon Blackburn’s “Just Causes” in his Essays in Quasi-Realism for a similar answer.
not, plausibly, have further causal effects in which we are interested. Compare: the fact that a liquid consists of HOH-molecules explains a great number of its properties and causal effects, but it is hard to see any additional effects stemming from the fact that its consisting of HOH-molecules makes it the case that the liquid is water. Or: the fact that a certain assembly of board, deal and batten was intended by its creator to be sat upon has certain effects (including the particular physical properties of the assembly), but the fact that this intention makes it the case that the assembly is a chair has no interesting additional effects. (This is not to say that the fact that a certain wrong/water/chair-making relation obtains has no effects, only that they are different from what we usually have in mind when we think of the effects of something being morally wrong, water, or a chair.)

Perhaps it might be thought that one such further effect is that we come to believe that the act is wrong. But it seems reasonable that what leads us to think that a certain act is morally right, or morally wrong, is that we recognize certain other facts about it, facts accessible by the ordinary perceptual and rational apparatus endowed by natural selection and training. If an action can be objectively right or wrong, the act’s moral rightness or wrongness is presumably not something that can be discerned independently of ordinary perception of ordinary properties.

On any account, then, it is those beliefs about an act that make us think that the act is morally wrong or right that are essential for our talk about moral rightness or wrongness in explanatory contexts. The difference between the objectivist and the non-representationalist account is the objectivist’s claim that the fact that an act has certain non-moral properties can make it true or false that the act is morally right or morally wrong independent of what we think about it, and not only make us think that it is morally right or wrong. That difference, however, will not importantly effect the way we would come to
give moral explanations: if the causal effects of the fact that an act is morally wrong are exhausted by the causal effects of its non-moral constituents or wrong-making characteristics, the fact that certain non-moral facts constitute or make certain moral facts hold will itself be without causal effects.

Admittedly, on any realist moral theory there must be some fact in virtue of which being an act of selfish theft makes that act wrong, and some fact in virtue of which the maker’s intentions makes an assembly of board, deal and batten into a chair rather than something that merely looks like a chair, and these facts will presumably have causal effects. For example, if an act is wrong for being a selfish theft because selfish thefts are paradigmatic cases of wrongful action, then that will have a number of causal upshots. It must be asked, though, whether it is these causal upshots that we have in mind when we take events to be causal effects of wrongful conduct. If it is not, then it presents no problem. If it is, on the other hand, the non-representationalist can admit that into the picture too: in saying that wrongful conduct caused the chairman’s downfall, we have in mind as causal agents (1) the fact that the chairman performed an act of selfish theft and (2) the fact that cases of selfish theft are among the paradigmatic cases of wrongful action, that is (approximately) among the acts that are central tests of a speaker’s proficiency with the word “wrong” as applied to acts in a moral context. In either case it seems that the non-representationalist can present an intelligible picture of making moral explanations.

So it appears that the non-representationalist can account for the fact that explanations in moral terms seem to be intelligible. And given that our moral opinions are based on representations of the acts they concern, it is only to be expected that if a person comes to think that her moral opinion has been caused in ways that have nothing to do with the act’s having properties whose representation has triggered her moral optation, she loses faith in that opinion. If there is something wrong with
non-representationalism, it is not that it cannot explain this tendency.¹

4.10. Invariance

There is a tree outside my window, and this fact does not essentially depend on whether I, or anyone, believes that there is a tree outside my window: if we all underwent some kind of brain-surgery or drug-therapy that somehow erased all our beliefs about trees, the tree could still be there. Similarly, it is (in all or most circumstances) morally wrong to take the life of an innocent person, and this does not essentially depend on whether I, or anyone, has the opinion that it is morally wrong. If we were brain-washed into having the opinion that such killing is right, it would still be wrong. At least that is how I and many others (though not everyone, certainly) tend to think about such counterfactual cases.²

Under representationalism, the fact that we have such thoughts seems wholly intelligible. They are thoughts about what constitutes moral wrongness and rightness or makes it the case that acts exemplify these properties, thoughts denying

¹ If someone were to establish objectivism, there would be a further task inviting explanatory moral judgements, namely to establish the truth of particular moral opinions. In the first part of Ethics, Inventing Right and Wrong, London: Penguin Books 1977, John Mackie argues that although moral opinions are (in a way) objective representations, we should believe that they are all false because there would be great explanatory costs, but no gains, in assuming the existence of the kind of facts that would make them true. A similar argument is put forth by Gilbert Harman in ch. 1 of his The Nature of Morality, New York, Oxford U. P. 1977. Both Mackie and Harman have provoked lively discussion. See for example Nicholas Sturgeon’s “Moral Explanations” in eds. David Copp and David Zimmerman, Morality, Reason, and Truth, Totowa N. J.: Rowman & Allenheld 1985; Geoffrey Sayre-McCord’s "Moral Theory and Explanatory Impotence" in Midwest Studies in Philosophy No. 12: “Realism and Antirealism”, Morris, Minn: University of Minnesota Press 1988; and Richard W. Miller’s Moral Differences, Princeton, N. J.: Princeton U. P. 1992, ch. 2.

² Of course, not all moral opinions come with a claim of invariance. For example, utilitarian opinions are grounded in beliefs about the consequences of actions for sentient well-being, and such consequences certainly vary with our desires and moral opinions.
that our thinking certain acts to be wrong has anything essential to do with it. But it might seem much more difficult to explain why we would have such thoughts if non-representationalism is correct and there are no facts “out there” which can make our moral opinions true or false independently of these opinions.

There is, however, a good explanation which follow the lines given in the last sections, and which was given in its essentials by Simon Blackburn in *Spreading the Word*.¹ It departs from the observation that the phenomena to explain are our *actual* moral opinions about the moral properties of actions performed in counterfactual cases where we lack these very opinions. When we have the opinion of certain actions that they are wrong and of other actions that they are right, these opinions are the reasoned results of believing that such actions have other, non-moral properties: that they have certain consequences on people’s welfare, express respect or disrespect for peoples’ autonomy, are acts of killing the innocent, and so forth. Since actions can have (and since we can believe that they have) those very same properties in the counterfactual case in which we have different opinions, it is only to be expected that we have the same opinions about these cases.

The point, then, is that the objects of our moral opinions (our moral grounds) are to a considerable degree independent of these very opinions. That is something which seems to hold for most of our opinions,² and perhaps most clearly for opinions concerning what will happen after our deaths. If we wanted our families (or theories, or favourite works of art) to fare well only as long as we wanted them to fare well, we would not want them to fare well after we are dead; but most of us do. (And it can indeed be argued that there is something incoherent in the thought of a desire that has *only* its own satisfaction as object.)

² Philip Pettit and Michael Smith defends this view in their “Backgrounding Desire”, pp. 565–592
It should be no surprise, then, that opinions consisting of optations elicited by moral grounds are accompanied by claims of invariance.

Still, our intuitions of invariance concern not only the rightness and wrongness of actions but also the validity of our moral opinions: it is not just that I think that it is wrong to kill the innocent, or that I think so on certain grounds, but also that it is right to think so on those grounds. As we saw in section 4.7, however, the non-representationalist can account for our opinions about what moral optations to have given certain facts or given certain beliefs, and for the fact that we might have reasons for such thoughts: such opinions are second-order optations about what moral optations to have under what circumstances, and their reasons are the facts or beliefs that elicited them. That these reasons obtain (if conceived of as facts) or are true (if conceived of as beliefs that such facts obtain) can be true or false quite independently of any thoughts that we have.\textsuperscript{1}

4.11. Concluding remarks

One popular argument for objectivism runs as follows: (1) Moral thought and discourse display the important features of an objectively representing practice – as the various arguments considered in this chapter have reminded us. (2) This means that any plausible version of non-representationalism must adopt all these features – as the non-representationalism considered in this chapter has done. (3) This, in turn, shows that the alleged non-representationalism is, after all, a version of objectivism.\textsuperscript{2}

\textsuperscript{1} For an extensive non-representationalist treatment of various claims of objectivity, I can heartily recommend Allan Gibbard's \textit{Wise Choices, Apt Feelings}, part III.

\textsuperscript{2} Crawford L. Elder argues that a non-representationalist theory must deny claims of invariance as discussed in section 4.10 or collapse into representationalism: see his "Antirealist and Realist Claims of Invariance" in \textit{The Southern Journal of Philosophy} Vol. 29, No. 1 1991. Paul Horwich makes a similar claim about Gibbard's non-representationalist theory in a review of \textit{Wise}
But such an argument would be premature until we have considered arguments against objectivism as it appears on that account: if we were to find a successful argument against objectivism, there would be something that stopped non-representationalism from collapsing into objectivism. Unfortunately, a convincing discussion of such arguments would demand us to be much clearer about what it is to represent an objective feature of the world. For example, is there any philosophically interesting sense in which the fact that a psychological state displays a number of representational-seeming characteristics makes it a representational state? If so, can we say what such a state represents? If Todd and Chip both have the opinion that it is morally wrong not to pay tithes, do their respective opinions (which consist of moral optations) represent the same fact? If so, in virtue of what? Any reader familiar with the extensive modern debate over the nature of representation will realise that it would take a full volume to start giving convincing answers to these questions, and will suspect that there might be several sensible senses of “represent” that would yield different verdicts. But since our subject in this essay is internalism rather than objectivism, there is no need to enter into such difficulties provided that we can supply the premisses of the argument of section 3.6. What we have thus far is:

(1) Moral opinions are, at least frequently, accompanied by the corresponding optations. (Sections 3.1–4.)

(2) At least in the good and strong-willed person, the considerations that elicit moral opinions – her moral rea-
sons – are those that elicit corresponding motivation.
(Section 3.5.)

(3) The externalist assumption that our moral psychology contains non-optational objective representations (“moral beliefs”, for short) adds nothing to our explanations of the phenomenology of the moral opinions of good people, or people that quite clearly have the relevant optations: the optations themselves account for the phenomena that moral beliefs could explain.

What remains to be shown in the next chapter is that:

(4) There are reasons to believe that the relevant optations obtain in most unclear cases too. Moreover, internalism can explain why we are inclined to think of certain cases as involving moral opinions even where it is clear that the relevant optations are lacking.

When (1) through (4) are established there will be no strong evidence that there are moral beliefs, but strong evidence that there are moral optations, and that moral optations are what we think of and identify when we think of and identify moral opinions.
5. Accommodation for amoralists

In chapter 3 and 4 we saw how a number of traditional arguments for and against internalism failed. The internalist arguments underestimated the externalist’s ability to explain plausible connections between moral opinion and motivation, while the arguments hoping to support externalism by vindicating objectivism underestimated the explanatory power of the internalist competitor. What the explanatory success of internalism suggests, however, is that externalism assumes more psychological states than needed to explain phenomena associated with moral opinions.

What we have, then, is the beginnings of an argument for internalism. But since that argument would be undermined by clear cases where we have moral opinions but lack the relevant optations, we shall now examine the claim that moral opinion and motivation come apart under certain circumstances, or that there are “amoralists”. It turns out, again, that proponents of these arguments have underestimated the accommodating powers of internalism.

5.1. The relevance of amoralism

If the argument of section 3.6 is correct, assertoric moral opinions are nothing but moral optations. But then there can be no cases of moral opinion unaccompanied by the relevant moral optation: no cases of amoralism. Hence if we can present a clear case of amoralism, the argument – and internalism – are undermined. Now, people have certainly thought that there are amoralists, and a variety of alleged cases have been presented. Before we discuss such cases in any detail, however, it is important to have a clear picture of their evidential connection to internalism.
Although the amoralist argument for externalism has been discussed in a number of recent articles, the most nuanced treatment of the amoralist character in a meta-ethical environment is probably found in Ronald D. Milo’s book *Immorality*. On the basis of his discussion of immorality, Milo puts forth an argument against internalism which starts with a claim about how we ordinarily conceive of wickedness:

...we think that the most evil or reprehensible kind of wrongdoing consists in willingly and intentionally doing something one believes to be morally wrong, either because one simply does not care that it is morally wrong [=amoralism\(^2\)] or because one prefers the pursuit of some other end to the avoidance of moral wrongdoing [=preferential wickedness]. Any analysis of the nature of moral beliefs that does not allow for even the possibility of these forms of immoral behaviour is, it seems to me, defective.\(^3\)

The problem with forms of internalism seems to be that they make these sorts of immorality impossible. Of course, our internalism is compatible with preferential wickedness, but it would rule out amoralism. If Milo is right, this is enough of a defect.

**Non-literal amoralism**  
In order to assess Milo’s claim, it is important to see in what way internalism *is* incompatible with the common sense phenomenology that Milo refers to. For

---


2 Milo’s use of “amoralism” differs from the present, and he uses “moral indifference” for what we have called “amoralism”.

3 *Immorality*, p. 253. Actually, Milo’s argument is explicitly an argument against non-cognitivism (such as Richard Hare’s), but its point is directed at the internalist character of non-cognitivist theories. Similar arguments are suggested in David O. Brink’s *Moral Realism and the Foundations of Ethics*, p. 27 and David Copp’s *Morality, Normativity and Society*, p. 16.
example, it is hard to believe that anyone in the internalist debate has ever wanted to deny that we can memorize and utter moral sentences, such as “it is morally wrong to eat meat” or “I must not hit my wife” and, following standard inferential patterns, draw consequences from such sentences without thereby having or expressing the corresponding optations. If the character that common sense is supposed to be familiar with and find reprehensible is a person able to recall the latter sentence but being in no way guided by it, internalists should not be worried: no sensible thinker should want to deny that there are psychological states resembling moral opinions which might be said to be moral opinions in an extended sense.

Compare with the state of being a Londoner. Suppose that John was born in Stockholm but lived in London from his seventh year until he was 22, when he went back with his family to Stockholm because he could not get the kind of job he was looking for in London. One month every summer he goes back to London, meets with his old friends and sees his old places. If someone would say that “He is a Londoner, exiled from London to his native Sweden by the unfavourable labour-market”, that would seem to make good sense (depending, of course, on a reasonable context). Plausibly, though, literally being a Londoner is to be a native or inhabitant of London. Take away that implication (“he is exiled to his native Sweden”) and the listener has to look for other aspects of being a Londoner that the utterance might have tried to communicate. We might, for example, suppose that the application of “Londoner” to John is meant to convey that London is where he feels at home, where he belongs. After all, home is where the heart is.

Similarly, if (1) literally being of the opinion that some act is wrong is to morally desire for that act not to be performed, and (2) having such a desire involves various tendencies to make inferences and utter, accept or deny sentences, then (3) the phrase “being of the opinion that some act is morally wrong without having the appropriate desires” might be comprehensible even if literally a contradiction in terms: having moral opta-
tions is to have a number of dispositions that can be displayed by other states of mind too.

It might be objected that this talk about the "literal" sense of an expression is misleading in implying that our concept of a moral opinion corresponds to a substance that can be defined in terms of sharp necessary and sufficient conditions, or in implying that we have sharp conventional rules of meaning for the application of expressions of the type "x is of the opinion that y is morally wrong." Rather, a psychological state is a moral opinion if it is a member of a set defined by a number of criteria that can be satisfied to degrees.¹

But internalists can insist that it is the fact that paradigmatic moral opinions consist of moral optations that explains why our numerous ways of recognizing moral opinions are what they are: moral optations are such as to get a reading along all the various dimensions used to detect moral opinions – in effect, this point was argued in chapter 4. (Similarly, the fact that paradigmatic samples of water consist almost entirely of HOH molecules explains why water has the kind of properties that we go by in identifying water, such as being liquid at room temperature, boiling at a certain temperature, reflecting light in a certain way, solving certain substances but not others, and so forth.) Given internalism, it is only to be expected that when we talk about people being in states resembling moral optations but lacking the relevant connection to action, this is sometimes done in terms of what they believe to be morally wrong, or what is morally wrong in their opinion – although it is equally to be expected that in order not to be misleading, such use of these expressions needs a context in which it is already obvious or

made obvious that the person lacks the relevant optations. (Similarly, other substances than water might behave as water, and it is not implausible that we would talk about such substances as “water”. Think of “fire-water” and heavy water consisting of DOD molecules.)

Moreover, it is only to be expected that there should be states resembling paradigmatic moral opinions, given (1) that we have an ability to reason from premisses that are not grounded in our own experience (or our own desires) but rather adopted from hearsay, (2) that some people lack paradigmatic moral opinions because they are insensitive to social and emotional clues and rarely if ever feel guilt, anxiety, sympathy or other feelings of importance in the development of moral optations,¹ and (3) that adopting and reasoning from the moral sentences of others can help a person who lacks moral optations to hide that sad and socially stigmatizing fact. So, from an internalist perspective, and holding platitudes about our capacities in mind, it seems quite intelligible that the amoralist should be part of our common sense way of thinking. Moreover, it does not seem odd that he causes moral outrage, since he threatens things that we find important.

Surprisingly, then, much of the phenomenology of amoralism is not only compatible with internalism but one of its natural consequences. What we have at this stage is the outlines of an explanation of why the nature of moral opinions has eluded externalists: both the argument for internalism presented at the end of chapter 4 and the explanation of the amoralist phenomenology have been based on considerations of greater methodological sophistication than is usual. But neither the argument nor the explanation has been completed: that will

take the remainder of this chapter, and sections 5.2–5 are devoted to rebutting various amoralist arguments in detail.

**Undermining internalism**  First, however, it needs to be shown that what I referred to as methodological sophistication does not amount to sophistry. Accepting the possibility of non-literal amoralists might seem to make internalism immune to counter-evidence, thus revoking its empirical status and banishing it to the realm of stipulation. If that was the case, the remaining central disagreement about internalism would seem to be a disagreement about words: whether to say that amoralists “really” have moral opinions. In fact, however, there are plenty of ways to discover the falsehood of internalism – if indeed it is false. It could have been that the internalist thesis failed the explanatory tasks in the last chapter, for example, and since it is a thesis about the functional organization of the processes and states that make up our moral thinking, possible upcoming problems lie in the future of neuropsychological research. Other possible ways to undermine internalism are more easily accessible today, however. For example, if we had everyday experiences of what we take to be, literally, moral opinions without moral optations, and if all possible internalist ways to account for such experiences were obviously *ad hoc*, that would be good evidence against internalism in proportion to how frequent and non-pathological the relevant cases are. (The reason that people are externalist when it comes to beliefs concerning such matters as the model of the Joneses’ new car or the colour of a pair of socks is presumably that we experience and expect all kinds of attitudes to come with such beliefs, depending on the context.)

We must be clear, though, about what the existence of such cases would show. Let us say that $M$ is the set of apparently clear-cut or literal cases of moral opinion, that $O^M$ is the subset of $M$ that involve the relevant optations, and that $A^M$ is the (amoralist) subset of $M$ that do not involve the relevant opta-
tions. Now suppose that AM has at least one actual\(^1\) member: what would that scenario suggest for internalism? If AM is very small, it would presumably suffice to say that there are a few cases of moral opinion that are not optations, but that internalism is — on the whole — a true account, as suggested by its explanatory success in OM. Moreover, if the members of AM are atypical in other ways than by being unaccompanied by moral optations, there would be reason for saying that they are not, literally, moral opinions: here, considerations of simplicity apply. If AM is of the same magnitude as OM, however, things would be different. Again, the internalist might insist that internalism holds for the members of OM, but that moral opinions have a “disjunctive essence”: they are either moral optations or ... (some other characterization of psychological states, perhaps a disjunctive characterization). Such is the case with jade, which consists either of jadeite or of nephrite, two quite distinct minerals.

But members of AM might undermine internalism not only by suggesting that it is at best true of a limited portion of moral opinions: they might suggest that internalism is false about members of OM, too. For example, suppose that all members of both AM and OM are analyzable as objective representations of moral reality. Then the fact that no member of AM consists of optation might suggest that optation is something that is added in cases that make up OM, and that an externalist account would be preferable: again, considerations of simplicity would apply. (At least if there are no other important differences between members of AM and OM apart from their different connections to optations, differences suggesting two distinct “essences”, as in the case of Jade.) If so, optations would seem

\(^1\) Note that our intuitions about hypothetical cases have little bearing on our object of study: the constituents of the actual psychological states that we identify as moral opinions, not of what could conceivably be thought of as moral opinions.
to be something that might or might not accompany moral opinions: something external, in other words.

What we shall do in the sections that follow is to look at everyday experiences of moral opinions that have seemed to be at odds with internalism, and see how well they can be accommodated in an internalist architecture without ad hoc parts. Room will be made for familiar phenomena such as moral weakness, the motivational role of moral experience, the possibility of having emotions disagreeing with one's moral opinions, and the acceptance of agent-relative moral obligations.

5.2. Moral weakness and listlessness

The existence and the interpretation of weakness of will and its relative moral weakness have been discussed at some length by philosophers, and apparent cases of moral weakness are sometimes invoked against internalist analyses of moral opinion.\(^1\) Looking closer at the issue, however, it is hard to see that such phenomena are relevant against internalism.

Suppose that we have a person who is, for the time being, morally weak. He is of the opinion that it is his moral duty to help his distressed friend that night. Falling suddenly and madly in love with a woman, however, he follows her home, leaving the friend to himself. Such cases are obviously problematic for internalist theories demanding that moral opinion should be followed by corresponding action.\(^2\) The present theory involves no such requirement, however: assertoric moral opinions are moral optations, but the connection between optations

---


2 In defending such a strong version of internalism, Richard Hare concludes that moral weakness is a matter of being unable to act according to one's moral judgement. See Freedom and Reason, pp. 77 ff.
and action can be broken in a number of ways – as stressed in sections 2.3 and 2.4.

A similar account is available for a related motivational malaise: listlessness. An agent struck by listlessness due to exhaustion, despondency or depression is not under the motivational conflict associated with weakness of will. Just like the morally weak, he can utter the normative phrases that he normally did before losing his list, but there is none of the other normal inner or outer signs of motivation to act accordingly. ("I know I ought to drink some water, but I just want to sleep"; "I know I should warn the others, but it means nothing to me: I just want to die") Once we see moral motivation as strategy-selecting states, however, we can insist that the listless agent nevertheless has an optation to act according to the moral opinion that he gives voice to. What listlessness does is not so much to change particular optations as to prevent whole strategy-selecting systems from affecting action, thereby blocking the agent’s moral opinion / optation from serving its proper function. Often, when that whole system is back in business, the particular optations that had ceased to affect action make themselves known more or less as they used to. Rather than assuming that they had been created anew it is reasonable to suppose that they were still there, although temporarily disabled or collectively half-cocked. And if that interpretation holds for most interesting cases of “amoralist” listlessness, such cases are no threat to internalism.

1 Michael Stocker argues against moral internalism with reference to listlessness on pp. 744–5 of his “Desiring the Bad: An Essay in Moral Psychology”. Unfortunately, this influential and important paper suffers from being fuzzy on two important points: it discusses the idea that the good is intrinsically attractive together with the idea that the believed good is intrinsically attractive, and it says far too little about what it is for something to attract.

2 As suggested by Aristotle two millennia ago, similar things can be said about those that are “asleep, drunk or mad”. See *The Nicomachean Ethics*, translated with introduction by David Ross, revised by J. L. Ackrill and J. O. Urmson, Oxford: Oxford U. P. 1980, Book 7.
5.3. Internalism and moral experience

In a recent argument against internalism, William Tolhurst urges us to reflect on the sources of moral opinions. To see why internalism is false, “we must consider the motivational significance of the bases on which a person can come to hold a moral belief; these can differ considerably.” Clearly, we can come to form moral opinions on the basis of personal experience, as well as on authority or reliable testimony, and Tolhurst is willing to say that internalism holds for moral opinions formed in response to experienced moral obligation:

To experience something to be obligatory or valuable is to have an experience that inclines one to believe it to be obligatory or valuable and that also tends to evoke attitudes that would incline one to act in accord with its having this normative status. Thus, when one comes to hold a normative belief on the basis of one’s own experience of value or obligation, one has an experience that is, under appropriate circumstances, capable of motivating one to act in accordance with the belief.

But moral beliefs formed on the basis of testimony are quite another matter, he claims. Here,

…the motivation, if any, must come from another source. … there may indeed be other motivational factors that will lead the person to perform the action judged to be obligatory. But again, there may not. This possibility is what makes room for moral insensitivity.

There seems to be something right in Tolhurst’s observation. Sometimes, people know – in some sense – that what they do is morally wrong, or that they ought, morally, to do something that they do not, but it is not until they really experience the

---

wrongness of what they do, or experience that they ought to do something, that they start acting the way they knew all along that they ought to act. Take the following case. Most of us know that in several areas on this earth people are dying en masse from starvation and disease that could be countered if the Red Cross or the Médecins Sans Frontières had more money, but we do nothing to help them. Suppose that one night we watch a television broadcast from one of these areas, representing in painful detail the misery of this world and the possibilities that are available for those interested in giving a helping hand. As a reaction, we decide to send some money (perhaps that day’s salary). We knew all along – really knew – that we could help, and many of us were of the opinion that we had a moral obligation to do so, but we did nothing until confronted with the suffering and death involved through the eye of the camera that we experienced or felt that obligation. The opinion in itself did not move us, only experience did.¹

Internalism would certainly be in trouble if it had no ways to (begin to) accommodate the varying motivational significance

¹ Talk about moral experience and its connection to moral motivation has been quite wide-spread in recent efforts to understand moral opinions. See for example John McDowell’s “Non-Cognitivism and Rule-Following”:

... we can learn to see the world in terms of some specific set of evaluative classification, aesthetic and moral, only because our affective and attitudinative propensities are such that we can be brought to care in appropriate ways about the things we learn to see as collected together by the classifications. (P. 142).

David McNaughton is sympathetic to McDowell’s account in his Moral Vision, e. g. ch. 7 and 8. Michael R. DePaul employs a visual vocabulary in arguing that moral knowledge not founded on moral experience is defective in “The Highest Moral Knowledge and the Truth Behind Internalism”, The Southern Journal of Philosophy 1990, Vol. XXIX, Supplement, and William Tolhurst shares DePaul’s view that the connection between moral experience and motivation is the truth of internalism; Alfred R. Mele considers the possibility that motivation is necessary for moral experience in his “Internalist Moral Cognitivism and Listlessness” and Lee Overton defends the same idea in his comments on that paper in Bears, 24/3 -97: http://www.brown.edu/Departments/Philosophy/bears/9703over.html.
of sources of moral opinion. But it has, or so I will try to show, starting with some general notes on moral experience.

_Moral experience_ Generally speaking, having or coming to have an opinion does not feel in any particular way. The thing about experiences, though, is that they do feel in particular ways: they are phenomenologically richer than mere opinions. If we ask in what way moral opinions can be phenomenologically rich—how we can “experience” wrongness or obligation—it seems that the phenomenology associated with optations is the most plausible candidate. On any view, the moral quality of an act does not look or smell or sound in any particular way, but moral optations often do have a phenomenology of sorts, giving rise to _emotional_ adjustments to the perceived situation. Such adjustments typically correspond to a readiness to perform a certain kind of act, to protect or destroy some object, or to display aggression (anger, indignation, resentment) or submission (guilt, shame), and might consist in changes in heart rate, skin colour, muscle tension, facial expression, posture, and so forth: such changes are typically phenomenologically rich. (See section 2.5.)

If I am right in taking this to be the story behind thoughts about moral experience, however, Tolhurst is putting the cart before the horse when he says that the experience is the _source_ of moral motivation. It seems much more plausible to say that moral experience is the experience of moral optations, or of being emotionally aroused by one’s moral grounds because one has certain moral optations. Of course, this objection to Tolhurst’s way of putting things does not show that his argument against internalism is flawed. Nevertheless, it does point to a way for the internalist to handle intuitions about the sources of moral motivation. Stronger optations—those that are more likely to lead to action “under the appropriate circumstances”—are more likely to result in emotions and in “moral experience”, and that might give rise to the illusion that moral opinions do not, by themselves, involve motivation, and that some kind of
(emotional) experience is required. In order to raise the credibility of this suggestion, however, we need details.

To begin with, the internalist could argue that in many cases where moral experience gives us motivation, it also gives us the corresponding moral opinion. We sometimes know that some action of ours has certain consequences for the welfare of other people, but fail to realise that these consequences make our action morally wrong until we get first-hand experiences of them. There are innumerable stories of soldiers who, after seeing their victims maimed and tormented, realize that it is morally bad to cause this sort of havoc. They certainly knew that shooting at people and bombing villages would cause death and suffering, but they did not include this when forming moral opinions until they had to look into the eyes or intestines of their victims. The reason for this might be that they used to think of the situation in terms or figures or images that had no optational impact, or that they had second-order optations keeping such considerations out of deliberation, optations that were changed by their experience.

Of course, not all cases are like that: we often have the opinion but not the kind of motivation that experience gives us. But internalism can accommodate that phenomenon, too, under various interpretations: having an optation does not imply feeling or acting in any particular way, and it seems that in typical cases where some previously existing moral opinion is activated by experience, that opinion was accompanied by some previously existing motivation. If we had been in the position to give the same amount of help as we came to give without taking most of the costs involved, we would have provided help even before having the motivating experience. Within an internalist archi-

---

1 This particular kind of experience can work the other way too, making the soldier habituated to killing. What the effects will be seems to depend very much upon the psychosocial milieu in which the experience is had.

2 For example, even if a person knows that smoking kills by cancer and that dying from cancer involves huge amounts of suffering, it might not be until she is confronted with the details that thinking about the risks affects her habits.
tecture there seem to be several ways to conceive of this increased motivation that are compatible with internalism: (1) Experience might increase the weight given in moral deliberation to the kind of suffering depicted in the television broadcast, thus increasing the moral strength of the opinion that it would be wrong not to help. (2) Experience might increase our confidence that the moral optation is not mistaken, thus increasing the doxastic strength of the opinion. (3) Experience might give us motivation beyond that resulting from moral deliberation: a personal desire to help more than we consider morally obligatory.

We should probably not suppose that these alternatives can or ought to be sharply distinguished, but there seem to be clear cases of all three, and taken together they surely accommodate much of what gave Tolhurst’s claim its strength. Being internalist interpretations of how experience increases moral motivation, however, they never (literally) allow the experience to be preceded by moral opinion without the relevant optation. Tolhurst’s point, of course, was exactly that moral opinions based on other things than relatively direct experience of the factors that gives an act its moral status might be unaccompanied by motivation. That is a point still to be substantiated, however, and from what we have seen this far there is no reason to expect this to happen.

5.4. Second-order desires and lagging moral optations
Suppose that our colleague at the department for applied ethics is reflecting on the possibility to genetically engineer “human” beings without brains in order to grow organs and tissue for transplantation purposes. He is initially appalled by the prospect, seeing it as one of the most abominable ideas ever suggested (right there with Swift’s “A Modest Proposal”). But thinking about the matter he decides that it would nevertheless be morally right to realize the idea, and indeed morally wrong not to. The gains for those in need of organs and tissue would outweigh the drawbacks, and the faint or nauseating feeling
evoked by visualizing the handling of these quasi-human bodies would be analogous to what medical students experience at their first vivisections: natural, but not morally “veridical”. As it happens, our colleague has a desire to bring his motivations and feelings in line with his reasoned moral conclusions, through contemplating the issues, undergoing therapy, or whatever it might take, but his moral optations does not yet seem to be in line with his new moral conviction. Can internalism account for this apparently realistic situation in plausible terms?1

Actually, there are two ways in which internalism can try to accommodate this and similar examples: by denying the conflict with internalism, and by denying that the example can be plausibly understood as involving a clear case of moral opinion. If we suppose that, in spite of negative emotional reactions, our colleague has moral optations that were elicited by his belief that cultivation of quasi-humans would serve the needs of those in line for transplants, then the example is in perfect consonance with internalism. The fact that the thought evokes negative feelings is simply irrelevant: it just reveals that he has negative optations too, but ones elicited by non-moral considerations or ruled out from deliberation by second-order desires.

If we suppose that our colleague lacks the relevant optations, on the other hand, the internalist must insist that he has no moral opinion, literally speaking. A person saying that it would be wrong not to cultivate quasi-humans on the ground that it would save human lives, but not at all desiring such a cultivation on those grounds could not comfortably be described as having the opinion that we ought to cultivate quasi-humans. (An externalist should admit this discomfort, but try to explain it along the lines of section 3.2.) However, since our colleague has a second-order desire to bring his motivation and feelings in line with his reasoned moral conclusions, he is in a complex

1 This problem for internalism was suggested by Torbjørn Tännsjö.
state of mind in some ways similar to a moral opinion and could therefore perhaps be described as having the moral opinion in an extended, non-literal, sense. He is not just saying and arguing that it would be morally wrong not to cultivate quasi-humans in organ factories; he is also in some ways acting the part. Lacking a better simple way to report that state of mind, we might say that he is of the opinion that it would be morally wrong not to cultivate quasi-humans, although he has no desire for it to be done.

It should be noted, though, that the state of mind described is quite strange. Though being very much aware that cultivating quasi-humans has positive consequences for human welfare — that was his ground for thinking that it would be morally wrong not to cultivate them — he has no desire for the act to be performed grounded on this awareness. Consider the following two alternative stories about his motivation:

1. He has no desire for human welfare to be promoted, but has a second-order desire to (i) desire the promotion of human welfare (because so desiring would make his desires consonant with the most simple and general moral theory, say) or to (ii) desire the realization of whatever his reasoning comes up with as “morally obligatory”. Neither suggestion has any obvious relevance for the nature of clear and paradigmatic cases of moral opinion.

2. Our colleague has a desire for human welfare to be promoted and likes the idea that organs would become available. These attitudes are transferred to a second-order desire to desire the action that promotes welfare

---

1 The latter alternative is especially strange since it would mean that he is motivated to be motivated by a conclusion the minor premise of which (“not setting up body-factories would miss out on positive consequences for human welfare”) failed to motivate. Here it would seem that all hinges on the expression “morally obligatory”. (The reader of section 3.5 will recall that Michael Smith charged the externalist conception of good people with turning such people into fetishists of this kind.)
and makes organs available, but they somehow fail to be transferred to a desire for the action itself. Again, this is something quite strange: if this kind of situation indeed occurs it is not obvious what it reveals about paradigmatic cases of moral opinion.

Hallvard Lillehammer describes a somewhat similar case in which a woman who is tired of her husband and temporarily indifferent to his feelings meets a charming man and is tempted to have an affair. She nevertheless comes to the conclusion that yielding to temptation would be wrong, “on account of her husband’s feelings”. Moreover, “a standing de dicto desire to do what is right” causes her not to cheat on her husband.1 If her moral opinion and her optation to perform the act have, as it seems, slightly different causal upshots (only the latter being elicited by a “desire to do what is right”), this would be a good counterexample to internalism if her state of mind is, as it seems, a perfectly ordinary case of moral opinion: if internalism is correct, her moral opinion is her moral optation and so necessarily has the same causal upshots. But when the details of the example are filled in so as to make it incompatible with internalism, it is no longer a striking everyday case of moral opinion, but – as we shall see – another case of strange word-fetishism.

Here are two ways of thinking about the example that make it look like a run-of-the-mill case of moral opinion but which also make it compatible with internalism. (A) The woman thinks that cheating on her husband is wrong because it would be a hurtful breach of confidence, and her “standing de dicto desire to do what is right” is in part a desire not to perform such acts. If so, however, she is motivated by the very consideration that elicited her moral opinion, even though her husband’s feelings as such fail to motivate. (B) The woman thinks that hurtfully breaking the confidence of her husband would be wrong because it would make her a cheater, and her desire to “do the right

thing” is in part a desire not to be a cheater. Again, the husband’s feelings are not by themselves what she takes to make the affair wrong, which means that the fact that they do not by themselves motivate her is compatible with an internalist interpretation.

Suppose instead that we discard these natural interpretations of the case and take seriously the suggestion that the woman’s thought that it would be wrong to have an affair was directly elicited by the thought that it would hurt her husband’s feelings. If so, we would face a person quite similar to the word-fetishist version of our colleague: she would be taking the very fact that cheating would hurt her husband’s feelings as making or constituting the wrongness of the act, but finding that fact motivating only when subsumed under the label of wrongness. That is hardly a run-of-the-mill example of moral opinion, and it is not in any way obvious what it tells us about more typical cases.

What we have seen in this section, then, is how seemingly normal and non-contrived cases seemingly at odds with internalism could be understood in various ways. Some of these ways yielded interpretations compatible with internalism while others turned the cases into something quite strange or psychologically implausible. If indeed such states exist, it is unclear what that reveals about the nature of paradigmatic cases of moral opinion.

5.5. Agent-relative obligations

If a theory about the nature of moral opinions made it unintelligible or very unlikely that people could have certain moral opinions when in fact many people do have these opinions, that would clearly tell against the theory. In this section, we shall see how internalism can handle the fact that people think there
are *agent-relative* obligations.¹ Consider the fact that some philosophers claim that

(1) there are agent-relative obligations, for example an equally strong obligation of every parent to favour the interests of his or her children over the interest of strangers

but at the same time claim that

(2) there is no obligation to see to it that others fulfill their obligations, no obligation, for example, to see to it that other parents take care of *their* children.²

Whether we should accept such theories or not is a matter of dispute, of course, but whatever the outcome of that dispute will be there seems to be nothing obviously or trivially incoherent in accepting them. If internalism is true, however, accepting such a theory might seem to be obviously incoherent. Accepting (1) amounts to having the moral optation that every parent favour the interests of his or her children over the interest of strangers, and that would seem to imply being (somewhat) motivated to see to it that every parent so favour his or her children. But accepting (2) implies *not* having a moral optation to see to it that every parent favour his or her children.

¹ In an argument similar to that discussed here, Thomas Nagel holds it against Richard Hare’s version of internalism that it rules out as logically misbegotten any theory that accepts agent-relative obligations. See his “Foundations of Impartiality” in eds Seanor, D. and Fotion, N., *Hare and Critics*, Oxford: Clarendon Press 1988. (I owe the existence of this paragraph to conversation with Michael Smith about his account of agent-relative reasons, as presented in “Internal Reasons” in *Philosophy and Phenomenological Research* Vol. 55, No. 1 1995.)

This could have been a *reductio* of internalism. It is not, however, since a careful application of our theory makes accepting (1) in conjunction with (2) quite intelligible. It is indeed true, on internalism, that accepting (1) implies having a moral optation that parents favour the interests of their own children. This, however, does not imply having a moral optation *to see to it* that other parents favour the interests of their children, and so does not imply having the moral opinion that one ought to see to that matter. The fact that an act is a case of favouring the interests of one’s own children is quite distinct from the fact that it is a case of making someone else favour the interest of her children: a person might focus on the former kind of facts but not the latter when forming her moral opinions. Consequently, internalists can say that there is no obvious “logical” incoherence involved.

But it might be thought that basic instrumental rationality – the uncontroversial core of practical rationality if there ever was such a core – guarantees that the optations of every rational person whose optations are elicited by the first kind of fact will be elicited by the second kind of fact too, and as strongly as by the first. If I deliberate rationally, have the optation that $p$, and realize that my doing $q$ would make it true that $p$, then I will presumably desire to do $q$, at least if there is no alternative way of making it true that $p$. So, if I am rational and have the moral optation that people favour the interests of their own children, then I would presumably have (the equally strong) desire (even if it is not a *moral* desire) to perform an act that makes it true that people favour the interests of their own children. That, however, seems to turn me into an awful busybody, as concerned with fulfilling my own obligations as seeing to it that others fulfil theirs. And yet it seems possible – clearly possible even – to have an agent-relative morality while being neither flagrantly irrational nor an aspiring moral missionary: hence it seems that internalism is amiss here.

A closer look will show that appearances are deceptive, however. First, I believe that accepting agent-relative obliga-
Commodations can and sometimes does turn a person into a moral busy-body – the phenomenon is not unheard of. But I also believe that there are ways in which we deny our optations their potential effects on our actions. One such way is by refusing to consider various means to the objects of our optations: refusing to bribe the umpire to guarantee the wished for victory of your favourite football team; refusing to lie in order to save a relationship; refusing to take one innocent life in order to save two; or refusing to neglect your own children to make sure that others take better care of theirs. Now, refusing to consider various means to a wished for or desired end is to employ a capacity that we mentioned in section 2.4, namely to choose to focus on and give weight to some consideration rather than another in one’s deliberation. And employing that capacity is not to be flagrantly or uncontroversially irrational. Rather, the capacity to choose what desires to act from has been seen as the crown of our rationality, most notably by Kant but more recently by writers such as Harry Frankfurt, who takes that capacity to be at the core of personhood,¹ and Christine Korsgaard who takes it to be a prerequisite for normative thought.² Moreover, even when the acceptance of agent-relative obligations is found on the sound side of the flagrantly irrational, it is not rationally unproblematic either: much of the debate about agent-relative obligations has concerned the rationality of such obligations.³

The conclusion, then, is that internalism renders agent-relative

³ To see how even proponents of agent-centred obligations are sensitive to these problems and see them as problems of rationality, see the introductory essay and texts by John Rawls, Robert Nozick, Philippa Foot, and Conrad Johnson in Samuel Scheffler’s (ed.) Consequentialism and its Critics; Thomas Nagel’s discussion of agent-relative reasons in chapter 9 of The View from Nowhere, New York: Oxford U. P. 1986, and Derek Parfit’s extended argument in Reasons and Persons, Oxford: Oxford U. P. 1984, that our reasons for action should become less agent-relative.
moralities psychologically possible but rationally somewhat problematic, and that this is just what they are.

5.6. The case for internalism

This chapter concludes our argument for internalism. We saw in chapter 3 that if we assume that most moral opinions are accompanied by moral optations of the relevant kind, we can explain that:

✓ People’s moral opinions vary with factors that tend to have an effect on their optations.

✓ Lack of moral optation is considered a sign of missing moral opinion.

✓ The typical object of trying to convince someone that some kind of act is wrong is to make that person desire or wish that such acts not be performed.

✓ Moral opinions are withdrawn when we learn that the act they concern could not be or could not have been performed.

✓ Moral judgements are apt for answering practical questions.

✓ When a good person changes her fundamental values, there is a reliable change in her moral optations.

To assume that *most* moral opinions are accompanied by moral optations is not to assume that *all* moral opinions are moral optations, of course. But it was argued in section 3.6 that internalism would be vindicated if it there is no explanatory need to assume that assertoric moral opinions are anything more than moral optations, and in chapter 4 we saw how it is possible to explain various striking objectivist features of our moral speech and thought without assuming that moral opinions are more than moral optations. In particular, it was argued that the
truth of internalism is consonant with or should lead us to expect that:

✓ Moral opinions are expressed by sentences in the indicative form.

✓ We use the standard patterns of inference and have the standard sort of sensitivity to inconsistency when reasoning in moral matters.

✓ Moral sentences are embedded in unasserted contexts, such as the antecedent of conditionals.

✓ We have permissive moral opinions.

✓ We feel that we can be mistaken in our moral opinions.

✓ We feel that we disagree rather than just differ in moral matters.

✓ We sometimes employ a moral vocabulary in explaining events.

✓ We feel that the moral status of many acts is independent of what moral opinions we have about them.

In this chapter, finally, it has been argued that the assumption that assertoric moral opinions are nothing but moral optations is consonant with a number of phenomena that has been supposed to be at odds with internalism, such as:

✓ Moral weakness and listlessness

✓ The motivating force of moral “experience”

✓ Lagging moral desires

✓ Seemingly indirect motivation

✓ Moral views involving agent-relative obligations

The amoralists that have haunted internalist theories are thereby accommodated within an internalist framework. This
does not mean that every instance of what could be somewhat plausibly described as (assertoric) moral opinion consists exclusively of moral optation: nature certainly provides some room for “freaks” and our fantasy is even more spacious. But freaky cases do little to undermine the conclusion that clear and paradigmatic cases of assertoric moral opinions consist of nothing but moral optations. In light of the phenomena considered, there is no need to make the externalist assumption that these cases involve further states representing moral facts.

Since the argument for internalism has been a kind of inference to the best explanation rather than a deduction from secure premisses, it may well be foiled by new hypotheses and a richer set of data. This is especially so since we have assumed without much argument that subjectivism is false, and since we have based the discussion on the assumption that some functionalist theory is true of the relevant psychological states. (Of course, subjectivism does not have the best of reputations, and functionalism is currently the most progressive programme in the philosophy of psychology, but such things can change.) Waiting for the jury of future research to reach its decision, however, internalism is a more promising hypothesis than externalism.
References

Altham, J. E. J.

Anscome, G. E. M.
• *Intention*, Oxford: Blackwell 1957

Aristotle

Arrington, Robert L.
• *Rationalism, Realism, and Relativism*, Ithaca, N. Y.: Cornell U. P. 1989

Axelrod, Robert

Baier, Kurt
• *The Moral Point of View*, Ithaca, N. Y.: Cornell U. P. 1958
• “Defining Morality without Prejudice” in *The Monist* Vol. 64, 1981, pp. 325–41

Bennett, Jonathan

Bergström, Lars
• *The Alternatives and Consequences of Action*, Stockholm: Almqvist & Wiksell 1966
• *Grundbok i Värdestori*, Stockholm: Thales 1990

Blackburn, Simon
• *Spreading the word*, Oxford: Oxford U. P. 1984

Brink, David O.
Campbell, Richmond
• “Critical notice of Wise Choices, Apt Feelings” in the Canadian Journal of Philosophy Vol. 23, No. 2 1993, pp. 299–324

Carlson, Erik
• Consequentialism Reconsidered, Dordrecht: Kluwer 1995

Copp, David

Damasio, Antonio
• Descarte’s Error, New York: G. P. Putnam’s Sons 1994

Dancy, Jonathan
• Moral Reasons, Oxford: Blackwell 1993

Davidson, Donald
• Inquiries into Truth and Interpretation, Oxford: Oxford U. P. 1984

Dawkins, Richard
• The Selfish Gene, Oxford: Oxford U. P. 1976

Dennett, Daniel C.
• The Intentional Stance, Cambridge, Mass.: The MIT Press / A Bradford Book 1987
• Consciousness Explained, London: Allen Lane 1991
• Darwin’s Dangerous Idea, New York: Simon & Schuster 1995

DePaul, Michael R.

de Waal, Frans

Elder, Crawford L.

Falk, W. D.
• “Ought and Motivation” in eds. Sellars and Hespers, Readings in Ethical Theory

Foot, Philippa
• “Moral Arguments” in Mind Vol. 67, 1958, pp. 502–13
REFERENCES

Frankena, William K.
- “The Principles of Morality” in ed. Goodpaster, Perspectives on Morality

Frankfurt, Harry G.

Gauthier, David

Geach, Peter

Gibbard, Allan
- “Reply to Sinnott-Armstrong” in Philosophical Studies Vol. 69, 1993, pp. 315–27

Goodpaster, K. E., ed.
- Perspectives on Morality, Notre Dame, Ind.: University of Notre Dame Press 1976

Gowans, Christopher

Grice, H. P.

Hare, R. M.

Harman, Gilbert
• The Nature of Morality, New York: Oxford U. P. 1977
• Change in View, Cambridge, Mass.: The MIT Press / A Bradford Book 1989

Hobbes, Thomas

Horwich, Paul
• “The Essence of Expressivism”, in Analysis Vol. 54, No. 1 1994, pp. 19–20

Hume, David

Hägglqvist, Sören
• Thought Experiments in Philosophy, Stockholm: Almqvist & Wiksell 1996

Kant, Immanuel

Keil, Frank C.

Kitcher, Philip

Korsgaard, Christine M.
• The Sources of Normativity, Cambridge: Cambridge U. P. 1996
Lillehammer, Hallvard
• “Smith on Moral Fetishism” in Analysis Vol. 57, No. 3 1997, pp. 187–95

Mackie, John L.

McDowell, John
• “Non-Cognitivism and Rule-Following” in Wittgenstein: to Follow a Rule, eds. Steven H. Holzman and Christopher M. Leich, Routledge and Kegan Paul 1981

McNaughton, David
• Moral Vision, Oxford: Blackwell 1988

Mealy, Linda

Mele, Alfred

Miller, Richard W.

Millikan, Ruth G.
• Language, Thought, and Other Biological Categories, Cambridge, Mass.: The MIT Press / A Bradford Book 1984
• “What Peter Thinks when He Hears Mary Speak” (commentary on Sperber and Wilson’s Relevance) in Behavioral and Brain Sciences Vol. 10, 1987, pp. 725–6
• White Queen Psychology and Other Essays for Alice, Cambridge, Mass.: The MIT Press / A Bradford Book 1993

Milo, Ronald D.
Moir, Anne and Jessel, David

Mortimore, Geoffrey W

Murdoch, Iris

Nagel, Thomas

Neander, Karen

Noggle, Robert
- “The Nature of Motivation (And Why it Matters Less to Ethics than One Might Think)”, *Philosophical Studies* Vol. 87, No. 1 1997, pp. 87–111

Nozick, Robert

Ohlsson, Ragnar

Overton, Lee
- Comment on Mele’s “Internalist moral cognitivism and listlessness” in *Bears*, 24/3 -97: http://www.brown.edu/Departments/Philosophy/bears/9703over.html

Parfit, Derek

Pettit, Philip and Smith, Michael

Plato
Price, Richard

Rescher, Nicholas
• The Logic of Commands, London: Routledge and Kegan Paul 1966

Sayre-McCord, Geoffrey

Scheffler, Samuel, ed.
• Consequentialism and Its Critics, Oxford: Oxford U. P. 1988

Schick, Frederic

Schueler, G. F.
• “Modus Ponens and Moral Realism” in Ethics Vol. 98, 1988, pp. 492–500

Searle, John R.
• Expression and Meaning, Cambridge: Cambridge U. P. 1979

Sellars, Wilfrid
• “Obligation and Motivation” in eds. Sellars and Hospers, Readings in Ethical Theory
• “On knowing the better and doing the worse” in International Philosophical Quarterly Vol. 10, 1970, pp. 5–19

Sellars, Wilfrid and Hospers, John eds.
• Readings in Ethical Theory, New York: Appleton-Century-Crofts 1952

Sinnott-Armstrong, Walter
• “Some problems for Gibbard’s norm-expressivism” in Philosophical Studies Vol. 69, 1993, pp. 297–313

Smith, M. B. E.
• “Indifference and Moral Acceptance” in American Philosophical Quarterly Vol 9, No. 1 1972, pp. 86–93

Smith, Michael
• The Moral Problem, Oxford: Blackwell 1994
• “Why Expressivist about Value Should Love Minimalism about Truth”, in *Analysis* Vol. 54, No. 1 1994, pp. 1–12

Solomon, David W.

Sperber, Dan and Wilson, Deirdre

Stevenson, Charles L.
• *Ethics and Language*, New Haven: Yale U. P. 1944

Stocker, Michael

Sturgeon, Nicholas

Tersman, Folke

Tolhurst, William

Tännsjö, Torbjörn
• *Moral Realism*, Savage, Maryland: Rowman & Littlefield 1990

Ullman-Margalit, Edna
• *The Emergence of Norms*, Oxford: Oxford U. P. 1977
REFERENCES

Warnock, G. J.
• *The Object of Morality*, London: Methuen 1971

Williams, Bernard
• *Problems of the Self*, London: Cambridge U. P. 1973

Wright, Crispin

Wright, Larry
• “Functions” in *The Philosophical Review* Vol. 82, April 1973, pp 139–168
Index

action-guiding function of moral opinions, as explaining connection to performability, 54
agent-relative obligations, 130–3
all-things-considered opinions, 83
Altham, J. E. J., 41n; 137
amoralism, 7, 112–34
• relevance of, 113–8
anger, 9; 31; 32; 36
Anscombe, G. E. M., 21n; 137
anti-rationalism, Humean, 13; 41
Aristotle, 121n; 137
Arrington, Robert L., 16n; 137
assertoric moral opinions, 38
Axelrod, Robert, 85n; 137
Baier, Kurt, 16n; 137
beliefs, 21; 24
• function of, to correspond to facts, 84
Bennett, Jonathan, 35n; 137
Bergström, Lars, 55n; 81n; 84n; 102–3; 137
Blackburn, Simon, 4; 25n; 58n; 74n; 77; 82n; 84n; 89n; 96n; 105n; 109; 137
blueprints, 82; 85; 92
Brink, David O., 4; 7n; 42n; 71n; 78n; 114n; 137
Campbell, Richmond, 7; 138
Carlson, Erik, 55n; 138
categorical moral opinions, 89
change in fundamental moral values and motivation, 61–72
codification, 64–5
coherece of desires, 70
command, 56–7
communication, 37
communicative function of moral sentences, 37
communicative point of utterances, 56
concepts, as upheld in response to fairly constant needs for communication and information processing, 54
conditional belief, as representing structures, 90
conditional moral opinions, 86–92
conflicting attitudes, 82
conflicting desire, 98
consistency, 80–6
• as mutual realizability, 82
cooperation,
• as evolutionary rationale for moral optations, 60–1n
• with a person with inconsistent optations, 85; 102
• with an agent lacking stable patterns of decision, 94
Copp, David, 27n; 36n; 87n; 114n; 138
copies, 6
correspondence with reality, 81; 90
Damasio, Antonio, 23n; 25n; 31; 33–4; 138
Dancy, Jonathan, 8n; 25; 41n; 49n; 65n; 138
Davidson, Donald, 21n; 25n; 138
Dawkins, Richard, 20n; 138
de dicto desire, 66; 129
de Waal, Frans, 32n; 61n; 138
Dennett, Daniel C., 20n; 25n; 64n; 138
DePaul, Michael R., 123n; 138
derived goals / desires, 67
desires (see also optations), 21; 24–5; 68
  • function of to cause facts to obtain, 21; 24–6; 85
  • de dicto, 66; 79; 129
  • derivative and non-derivative, 66–73
  • instantaneously created, 67
  • coherence of, 70
  • as behaving similarly to beliefs, 76
  • clashes of, 83
disagreement in moral opinions, 43–47; 94–102
  • as compared to disagreement in non-moral factual matters, 46
  • as dissolved by means of argument, 94–6
disjunctive moral opinions, 87
dispositions to form desires, 67–8
doxastic strength, 104
Elder, Crawford L., 110n; 138
emotions,
  • moral, as connected to moral opinions, 31–36
  • moral, as not being connected to moral opinions by optations, 36
  • as fulfilling cognitive functions, 34
  • functions of, 31–2; 34; 36; 61n
“ethical judgement”, 13
explanations, moral, 104–8
  • as essential to confidence in our moral views, 105
expressing a moral opinion, 37
external desires (1), 71
external desires (2), 72
externalism, 7; 9; 41–2
  • as the negation of internalism, 39
  • Michael Smith’s theory of moral opinions as a species of, 69
Falk, W. D., 5; 13n; 138
fallibibility, intuitions of, 100
fear, 31; 33
fetishism, 66; 69; 71; 129–30
folk-psychology, 19
Foot, Philippa, 16n; 131n; 133n; 138
Frankena, William K., 5n; 16n; 139
Frankfurt, Harry G., 133; 139
Frege, 87
functional organization of our psychological apparatus, speculation about, 77
functions, 219–20
  • as the most important thing something actually does, 58
  • as what a thing does most often, 59
  • normative, 59
  • of beliefs to correspond to facts, 20–1; 24–5; 84
  • of optations to cause facts to obtain, 20–1; 24–5; 85
  • of emotions to prepare for action, 31–4
  • of moral emotions to affect the action of others, 36
fundamental (moral) values, 63–5
Gauthier, David, 61n; 71n; 139
Geach, Peter, 87; 139
Gibbard, Allan, 4; 8n; 17–8; 25n; 35–6; 61n; 75n; 77; 82n; 85n; 110n; 120n; 139
good people, 61–74
goodness,
  • as fetishistic on externalist theories, 66
• and strength of will, 62; 64; 71
Goodpaster, K. E., 139
Gowans, Christopher, 139
Grice, H. P., 57n; 139
guilt, 31–2
Hare, R. M., 5; 31n; 52; 55–6; 60; 76; 87n; 120n; 131n; 140
Harman, Gilbert, 83n; 108n; 140
Hobbes, Thomas, 61n; 140
Horwich, Paul, 110–1n; 140
Hospers, John, 143
Hume, David, 5–6; 25; 140
• anti-rationalism of, 13; 41
hypothetical moral opinions, 38
Håggqvist, Sören, 8n; 140
imagination, as evidence for / against internalism, 7–8
impartiality, 27
improvement of moral opinions, 95–6
inconsistency,
• of blueprints, 82
• limited tendency to shun, 83
• representations, 95
• as implying mistakes, 95
• responsiveness to, 80–6
indicative mood, 78–9; 87
• as suited for inferences and conditional statements, 79
indifference, as source of perplexity, sign of insincerity, and sign of failed argument, 48–51
indignation, 31
inference, 79–94
insincerity, as indicated by indifference, 51
intentions, 56; 57n; 83
internalism, 3; 12–39
• historical survey of, 3
• roughly defined, 12
• the case for, 134
• as a theory about metaphysical necessity, 12
• as a conceptual truth, 11
• as the result of stipulation, 12
• as an empirical hypothesis, 11–3; 18–9; 77
introspective access to strategy-selecting states, 23
irrationality (see also rationality), 53; 89; 132–4
Jessel, David, 45n; 142
Johnson, Conrad, 133
“judgement of duty”, 13
“judgement of obligation”, 13
Kant, Immanuel, 4; 7; 133; 140
Keil, Frank C., 116n; 140
Kitcher, Philip, 61n; 140
Korsgaard, Christine M., 133; 140
language of morals, 36
Lillehammer, Hallvard, 141
listlessness, 120–1
“literal” sense of an expression, 116
Mackie, John L., 108n; 141
Malmnäs, Per-Erik, 9n
malfunction, 21
mapping, 20; 81–2
McDowell, John, 64–5n; 123n; 141
McNaughton, David, 65n; 120n; 123n; 141
Mealy, Linda, 117n; 141
Mele, Alfred, 8n; 123n; 141
Miller, Richard W., 108n; 141
Millikan, Ruth G., 20n; 25n; 41n; 58n; 83n; 116n; 141
Milo, Ronald D., 16n; 114; 141
minimalism, 111n
mistakes, 94–102
Moir, Anne, 45n; 142
Moore, G. E., 75n
moral causes, 104
moral clashes, 83
moral considerations, 18
moral deliberation, see practical deliberation
moral emotions, 32–4
  • and moral opinions, 31–6
  • function of, to affect the action of others, 36
moral experience, 122; 124
moral explanations, 104–8
moral internalism, see internalism
“moral judgement”, 13; 56
moral opinions, 13–14
  • means of identifying, 14
  • identification of, as going beyond identification of the relevant vocabulary, 15
  • identification of, as involving identification of the considerations on which the opinion is based, 16–8
  • as identified with the results of practical deliberation concerned with moral considerations, 29
  • assertoric, 38
  • permissive, 38
  • hypothetical, 38
  • categorical, 88
  • as based on beliefs about epistemically problematic matters, 46
  • improvement of, 95
  • about counterfactual cases, 108
  • doxastic strength of, 104
moral optations, 27–8
  • interest in avoiding inconsistent of, 82–6
  • as connected to moral emotions, 34
  • as having objects that are highly independent of moral opinions, 110
moral outlook, 64–5
moral properties, see moral rightness or wrongness
moral reality, structure of, 90
moral rightness or wrongness,
  • nature of, 50; 53
  • as not discerned independently of ordinary perception of ordinary properties, 108
  • as independent of moral opinions, 108
moral sentences, communicative function of, 37
moral strength (of moral opinions), 104
“moral thought”, 13
moral vision, 64
Mortimore, Geoffrey W., 142
motivation, 19–27
  • models of, 9
  • and motives, 24
Murdoch, Iris, 30n; 64n; 142
Nagel, Thomas, 41n; 131n; 133n; 142
Neander, Karen, 20n; 142
Noggle, Robert, 25n; 142
non-derivative desires (for moral values), 66–72
Non-literal amoralism, 114
non-moral constituents of moral facts, 107
non-representational, 75
normative function, 59
“normative judgement”, 13
Nozick, Robert, 68; 133n; 142
o-features, 45
o-relativity, 45
objectivism, 9; 41–3; 73; 75–112; 119
Ohlsson, Ragnar, 83n; 142
optational consistency, 85
optational strength, 103
optations (see also desires), 9; 28
• as being responsive to the world, 76
• function of, to cause facts to obtain, 85
“ought” imply “can”, 51
“ought”-judgement”, 13
Overton, Lee, 123n
paradigmatic cases of moral opinion, 14
Parfit, Derek, 133n; 142
particularism, 64
passions, as being beyond rational criticism, 6
performability, 53
permissive moral opinions, 38
• and consistency, 92
Pettit, Philip, 109n; 142
phenomenological access to strategy-selecting states, 23 plans, 82
Plato, 61n; 142
practical deliberation, 29–31
• by collectives, 93
• rational, 132
practical judgement, 9
practical nature of morality, 47–51; 53–55
• internalist explanations of, 49
• externalist explanations of, 50
practical questions, 57
pretence and acting, 15
Price, Richard, 3; 143
rational criticism, 6; 39–40; 70
rationality, 71; 133–4
Rawls, John, 133n
reason, as “the slave of the passions, 5
relativism, 43; 101
representationalism (see also objectivism), 6; 40
representing, 20
Rescher, Nicholas, 82n; 143
resentment, 31; 124
Sayre-McCord, Geoffrey, 108n; 143
Scheffler, Samuel, 143
Schick, Frederic, 30n; 143
Schueler, G. F., 82; 143
Searle, John R., 57n; 87n; 143
second-order optations, 35; 92; 98; 102; 110; 125–9
Sellars, Wilfrid, 13n; 143
Sinnott-Armstrong, Walter, 8n; 84; 143
Smith, M. B. E., 49n; 143
Smith, Michael, 4; 16n; 25n; 26n; 49n; 61–71; 109n; 111n; 128n; 143–4
Solomon, David W., 114n; 144
somatic markers, 33n
somatosensory cortex, 33–4
Sperber, Dan, 37n; 57n; 144
Stevenson, Charles L., 5; 13n; 37; 43–4; 48; 76; 96n; 100n; 144
Stocker, Michael, 7n; 121n; 144
strategy-selecting states, 19–24; 33
• conflict of, 22
• epistemic access to, 22
strategy-selecting systems, prevented from affecting action, 121
Sturgeon, Nicholas, 75n; 104; 108n; 144
subjectivism, 40–1; 78; 101; 136
Tersman, Folke, 82n; 144
Tolhurst, William, 122; 123n; 124; 126; 144
truth, 6; 40
<table>
<thead>
<tr>
<th>Author</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tännsjö, Torbjörn</td>
<td>81n; 127n; 144</td>
</tr>
<tr>
<td>Ullman-Margalit, Edna</td>
<td>61n; 85n; 144</td>
</tr>
<tr>
<td>unasserted contexts</td>
<td>86–92</td>
</tr>
<tr>
<td>uncertainty</td>
<td>102–4</td>
</tr>
<tr>
<td>unstable deliberation</td>
<td>as undermining long-term planning, 94</td>
</tr>
<tr>
<td>valid inference</td>
<td>87; 88</td>
</tr>
<tr>
<td>&quot;value judgement&quot;</td>
<td>13</td>
</tr>
<tr>
<td>voluntary action</td>
<td>24</td>
</tr>
<tr>
<td>Warnock, G. J.</td>
<td>16n; 145</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>weakness of will, 35; 120n</td>
</tr>
<tr>
<td></td>
<td>wickedness, 114</td>
</tr>
<tr>
<td></td>
<td>Williams, Bernard, 86n; 131n; 145</td>
</tr>
<tr>
<td></td>
<td>Wilson, Deirdre, 37n; 57n; 144</td>
</tr>
<tr>
<td></td>
<td>wishes, as connected to desires, 27–8</td>
</tr>
<tr>
<td></td>
<td>Wittgenstein, Ludwig, 65n</td>
</tr>
<tr>
<td></td>
<td>Wright, Crispin, 111n; 145</td>
</tr>
<tr>
<td></td>
<td>Wright, Larry, 20; 145</td>
</tr>
<tr>
<td></td>
<td>wrong-making characteristics, 107</td>
</tr>
</tbody>
</table>