

OXFORD

# PLATO'S MORAL PSYCHOLOGY

Intellectualism, the Divided Soul,  
and the Desire for Good



RACHANA KAMTEKAR

# Plato's Moral Psychology



# Plato's Moral Psychology

*Intellectualism, the Divided Soul, and the Desire for Good*

Rachana Kamtekar

OXFORD  
UNIVERSITY PRESS

# OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,  
United Kingdom

Oxford University Press is a department of the University of Oxford.  
It furthers the University's objective of excellence in research, scholarship,  
and education by publishing worldwide. Oxford is a registered trade mark of  
Oxford University Press in the UK and in certain other countries

© Rachana Kamtekar 2017

The moral rights of the author have been asserted

First Edition published in 2017

Impression: 1

All rights reserved. No part of this publication may be reproduced, stored in  
a retrieval system, or transmitted, in any form or by any means, without the  
prior permission in writing of Oxford University Press, or as expressly permitted  
by law, by licence or under terms agreed with the appropriate reprographics  
rights organization. Enquiries concerning reproduction outside the scope of the  
above should be sent to the Rights Department, Oxford University Press, at the  
address above

You must not circulate this work in any other form  
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press  
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data  
Data available

Library of Congress Control Number: 2017946738

ISBN 978-0-19-879844-6

Printed and bound by  
CPI Group (UK) Ltd, Croydon, CR0 4YY

*To my mother, Gurdip Kamtekar, and the memory of my father, Dileep Shankarrao Kamtekar (18 December 1928–25 November 2013).*



# Contents

<i>Acknowledgements</i>	ix
Introduction	1
1. Doctrine and Dialectic in Plato's Dialogues	6
2. Psychology for Sophists	35
3. Why is Wrongdoing Unwilling?	69
4. The Divided Soul	129
5. Why is the Divided Soul <i>Tripartite</i> ?	165
6. Psychological Eudaemonism and Explanation	186
<i>Bibliography</i>	209
<i>Author Index</i>	219
<i>Index of Passages</i>	221
<i>Subject Index</i>	229





# Acknowledgements

I am lucky in philosopher friends, and so have many people to thank for their help on my book.

Rachel Singpurwalla and Michelle Kosch read the whole manuscript and gave me written chapter-by-chapter comments, as did two anonymous readers for OUP. Earlier, Clerk Shaw read and commented on Chapters 1–3. Thank you all so much!

In addition to talking through many of the ideas in this book with me and commenting on written versions of most of the chapters below over many years, in June 2015 Stephen Menn organized a workshop around the manuscript at the Humboldt University, Berlin. I'm grateful to him and the other participants—Alison Laywine, Hakan Genc, Ronja Hildebrandt, David Merry, Christopher Roser, and Antonio Vargas—for their comments on Chapters 1–3 and 6.

Manidipa Sen organized a series of seminars on Plato's moral psychology at Jawaharlal Nehru University, New Delhi, in August 2016 during which I presented the material in Chapters 1–5 of the book, and I'm thankful to the faculty and students who participated, especially Kranti Saran and Soham Shiva, for their questions and comments.

Early instalments of much of the material in the book were presented to students in my seminars at the University of Arizona in spring 2013 and Cornell University in spring 2014. I am grateful to the students (especially Jeremy Reid, who later read the whole manuscript), and to Julia Annas, Tad Brennan, and Charles Brittain, who sat in on these seminars, for their thoughts about material I presented. Some of the material in Chapters 3 and 4 descends from a seminar I gave at the University of Chicago in spring 2008, and I would like to thank the students in that seminar, as well as Gabriel Richardson Lear and Jonathan Lear, who sat in, for their pointed questions about the moral psychology of the *Republic*.

I'm grateful to André Laks, Fernando Muñoz, John Proios, and an audience at the Facultad de Filosofía at Universidad Autónoma Nacional de México for their comments on previous versions of Chapter 1. Philosophy Colloquium audiences at the University of Wisconsin-Madison, Harvard

University, Mount Holyoke College, Washington University, and the University of Pittsburgh heard the material in Chapter 2 and their questions helped me clarify my ideas; Rachel Barney, Tad Brennan, and Agnes Callard also gave me valuable detailed comments on this chapter. Philosophy Colloquium audiences at Carleton College, St Olaf College, Davidson College, Georgetown University, Duke University, the Ancient Philosophy Workshop at the University of Chicago, and the University of California Riverside workshop on Ancient Moral Psychology all heard or read parts of Chapter 3. Thanks to those audiences for their help as well. And thanks to Shaun Nichols for helpful discussion about the material in Chapters 3 and 6.

Chris Bobonich first sparked my interest in Plato's moral psychology, taught me, and gave me many opportunities to develop my ideas on the topic, encouraging me throughout. I couldn't have asked for a better advisor and friend.

John Proios provided invaluable last-minute assistance compiling the indexes to the book. I'm of course responsible for any errors.

Finally, I am grateful to the Philosophy Department at the University of Arizona for periodically giving me time off to work on this book over the many years it's taken me to finish it, and to Terry Irwin and Gail Fine for sharing their library study and home with me when I needed a place to work in Ithaca.

There is one non-philosopher without whose unfailing love and encouragement this book could not have been written: Satwik Kamtekar, thank you.

# Introduction

If we want to know how we might live well, or live a good life, it is natural to ask who *we* are (*Alcibiades I* 128e–129a), and in particular what the capabilities and limitations are that we as human beings and as individuals bring to living our lives. Famously, Plato's Socrates identifies that by which we live well or badly, and by which we are just or unjust (*Crito* 47e–48a), with that by which we live, that is, the soul (*psychê*). So by 'Plato's moral psychology' I understand Plato's account of the soul insofar as it is relevant to our living well or badly, virtuously or viciously. While Plato's moral psychology ranges more widely than this book, taking in topics that I cannot do justice to here, such as the nature of pleasure and the relationship between our souls and the soul of the world, I focus on the core of Plato's moral psychology, human motivation.

This book argues for a new account of Plato's thinking about human motivation across the dialogues. A new account is needed because the dialogues do not seem to present a consistent account of human motivation, and existing accounts of how Plato's views about human motivation developed are not convincing. In Plato's early dialogues, Socrates seems to maintain that

- (1) virtue is knowledge (e.g. *Protagoras* 360c–d; *Laches* 192c–d) and vice ignorance;
- (2) wrongdoing is involuntary (e.g. *Gorgias* 509e);
- (3) we always do what we believe is the best of the things we can do (*Protagoras* 358c).

According to the mainstream account of Plato's moral psychology,<sup>1</sup> (1), (2), and (3) belong to a 'Socratic intellectualist' package, with (3) being the theoretical basis for (1) and (2), and implying that

- (4) there are no non-rational or good-independent motivations.

On this interpretation, Plato rejects (3) and (4) with the introduction of the divided soul in the *Republic* on the grounds that they are incompatible with certain phenomena of psychic conflict, such as action contrary to one's belief or knowledge about what's best, and that they cannot account for the behaviour of pre-rational human children or non-rational animals.<sup>2</sup> Instead of (3) and (4), Plato's middle and late dialogues suppose the soul to be divided into rational or good-directed and non-rational or good-indifferent parts, each independently capable of moving us to action, sometimes contrary to what we know or believe to be best.

There are several problems with the mainstream account. First, the dialogues that reject (4) and acknowledge that we have motivations contrary to our judgement of what's best (such as thirst or the desire for drink even when we know that the drink is unhealthy for us), and arguably also reject (3), countenancing akratic action, nevertheless maintain (2), that no one is willingly bad (this is the case in the *Republic*, *Timaeus*, and *Laws*). But if (3) is the theoretical basis for (2), on what grounds can these dialogues reject (3) but uphold (2)? Second, in the *Protagoras* Socrates introduces (3) after considering a case of apparent akrasia in which one's knowledge of what is best seems to be overcome by the pleasure of something else, and shows how (3) can explain it, so how can this same phenomenon be his grounds for rejecting (3) in the middle dialogues? Third, in several early dialogues where Socrates is supposed to hold (3), he seems to deny (4) and to recognize that we have desires for things other than what is best for us. For example, in the *Charmides* Socrates contrasts appetite, which is for pleasure, with wish, which is for a good (167e). What is to stop such desires from motivating us to act?

Recently some scholars have proposed alternatives to the mainstream account, arguing (a) that Plato, or the Socrates of the middle dialogues, also denies that non-rational motivations generate action without the mediation of reason (thus continuing to maintain (3) and (4)),<sup>3</sup> or (b) that the Socrates of the early dialogues also recognizes non-rational motivations but holds that they motivate action only indirectly, by influencing the agent's beliefs (thus rejecting (4) and complicating (3)).<sup>4</sup>

However, the idea (a) that Plato retains (3) across the dialogues is difficult to square with his clear rejection of (4) and recognition that there are motivations that can operate independently of one's reasoned judgement in the *Republic*, *Timaeus*, *Phaedrus*, and *Laws*. And the idea (b)

that even the early dialogues reject (3), so that the difference between them and the middle dialogues is whether non-rational motivations cause our actions only via the mediation of reason or also directly, leaves it mysterious what philosophical reasons Plato could have for such a shift.

The main thesis of this book is that rather than (3), human beings only do what we believe to be the best of the things we can do, it is

(3\*) human beings have a natural desire for our own good

that is the moral psychological foundation for both Socratic intellectualism and the divided soul of the middle dialogues. This natural desire for our own good may be manifested in different ways: certainly by our pursuit of what we believe to be best, but also by our pursuit of pleasant things and fine things.

Here is a very brief outline of the argument to come. Chapter 1 lays out the methodological approach employed throughout the book, which is to pay attention to the dialectical dependence of what the main speaker in the dialogue says on the intellectual problem(s) set up in the dialogue both by himself and the other speakers. To illustrate, Chapter 1 describes Socrates' use of hypotheses to answer questions that go beyond his claims to knowledge in the *Republic*.

Chapter 2 argues that in the *Protagoras*, Socrates hypothesizes (1) 'virtue is knowledge and vice ignorance' because if true, it would explain how virtue can be taught (as Protagoras claims), and then argues for a 'higher' hypothesis, (3) 'we always do what we believe to be the best of our options', on the basis of a 'highest' hypothesis, 'pleasure is the good', for if true, these higher hypotheses would explain how virtue can be knowledge (as virtue's teachability seems to require). The identification of the good with pleasure serves not only to introduce (3) in the *Protagoras* but also to replace a popular conception of the agent as moved to act by the strongest of competing forces with a conception of the agent as a subject representing actions as good or better and bad or worse and acting on what appears best. However, the dependence of (3) on ethical hedonism in the *Protagoras* should caution us from exporting (3) to other dialogues, since in every other dialogue, Plato's Socrates argues that pleasure is not the good.

Chapter 3 argues that (2) and various related expressions across the dialogues—'no one does *wrong* willingly' (*Gorgias*), 'no one does *bad* or shameful things willingly' (*Protagoras*), and 'everyone is bad or vicious unwillingly' (*Republic*, *Timaeus*, and *Laws*)—are based not on

(3) but on (3\*). (2) generalizes an argument-form used by (roughly) contemporary intellectuals to excuse agents from blame for bad actions—namely, that the agent was overwhelmed or compelled by some force (such as, for example, a god, or passion), for an agent would have to be compelled to pursue what is contrary to their natural desire for their (real) good (3\*). Like his contemporaries, Plato treats contrariety to the agent's natural desire for good as grounds for calling an action or state unwilling. As for ignorance, it is a condition that makes compulsion possible. In dialogues before the *Republic*, Plato also uses contrariety to the agent's natural desire for good to 'deattribute' desires or beliefs from an agent, despite that agent's avowals. (This isn't the same as denying that there are any good-independent motivations; rather, it is disowning them.)

However, unlike his contemporaries, Plato develops (3\*) by giving a teleological account of a complex human soul. If the argument of Chapters 2–3 is correct, soul-division can't be for the sake of rejecting (3) to acknowledge the existence and independent motivational efficacy of good-indifferent desires as on the mainstream interpretation. Instead, Chapter 4 argues, the point of soul-division in the *Republic*, *Phaedrus*, and *Timaeus* is to recognize human nature's multiple and potentially conflicting natural orientations, an upshot of which is to block deattribution of motivations, even 'unwilling' and self-destructive ones. Nevertheless, none of these motivations are good-indifferent, insofar as they are constructed to be sensitive to some aspect of goodness. Chapter 4 argues that this good-sensitivity is captured in the *Republic*'s characterization of soul-parts as homuncular sub-agents, each having a more or less limited conception of goodness—a characterization of soul-parts that also enables us to manage our own conflicting motivations as we aim at virtue.

Chapter 5 is concerned with *tripartition*—with why Plato insists that there is, in addition to the reasoning and appetitive parts of the soul, a third, 'spirited' part. I argue that by isolating a spirited part of the soul that is 'by nature a helper' of the reasoning part, Plato is able to distinguish a conception of reason as seeking theoretical knowledge from a conception of reason as belief in conformity with law and calculation; this distinction is the basis for his conception of philosophical virtue as distinct from ordinary virtue.

Finally, Chapter 6 discusses the relationship between these claims about Plato's account of human motivation and psychological explanation, arguing that (3) does not by itself provide sufficient explanations for actions. While an intelligent agent's beliefs about what is good explain her actions, an unintelligent agent's false beliefs themselves call for explanation.

## Notes

1. e.g. T. H. Irwin, *Plato's Moral Theory* (Oxford, 1977), 78–80; T. H. Irwin, *Plato's Ethics* (Oxford, 1995), 52–5, 116–17, 138–9; M. Frede, 'Introduction', in S. Lombardo and K. Bell (trans.), *Plato: Protagoras* (Indianapolis, 1992) [= 'Introduction'], xxix–xxx.
2. e.g. Irwin, *Plato's Moral Theory*; C. D. C. Reeve, *Philosopher-Kings: The Argument of Plato's Republic* (Princeton, 1988) [= *Philosopher Kings*], 134–5; T. Penner, 'Plato and Davidson on the Parts of the Soul and Weakness of the Will', *Canadian Journal of Philosophy*, 16 (supplement) (1990) [= 'Plato and Davidson on the Parts of the Soul'], 35–74.
3. e.g. G. Carone, 'Akrasia in the *Protagoras* and *Republic*: Did Plato Change His Mind?', *Oxford Studies in Ancient Philosophy*, 20 (2001) [= 'Akrasia in the *Protagoras* and *Republic*'], 107–28. Less committal, but in roughly the same camp is J. Wilburn, 'Akrasia and the Rule of Appetite in Plato's *Protagoras* and *Republic*', *Journal of Ancient Philosophy*, 8 (2014) [= 'Akrasia and the Rule of Appetite'], 57–91.
4. D. Devereux argues that Socrates recognizes the existence of non-rational motivations but considers only rational motivations relevant to virtue in 'Socrates' Kantian Conception of Virtue', *Journal of the History of Philosophy*, 33 (1995), 381–408; T. Brickhouse and N. Smith argue that Socrates accords non-rational desires an indirect role in motivation (they cause appearances which influence beliefs about good and bad which determine actions) in *Socratic Moral Psychology* (Cambridge, 2010) [= *Socratic Moral Psychology*]. I will not be discussing the view of H. Segvic that Socratic knowledge includes the non-cognitive elements ordinarily thought to be necessary for virtue: 'Certain desires and feelings are part of the knowledge that is virtue... The virtuous person can do what he wants to do because the taking-to-be-good that his willing amounts to is itself a state of knowledge: it is an accurate grasp of what is in fact good', in 'No-One Errs Willingly: The Meaning of Socratic Intellectualism', *Oxford Studies in Ancient Philosophy*, 19 (2000) [= 'No-One Errs Willingly'], 1–45 at 39.



# 1

## Doctrine and Dialectic in Plato's Dialogues

### 1.1 Introduction

Historians of philosophy aim to understand what historical philosophers thought about various topics of philosophical interest, and why they thought these things—rather than focusing, as might a non-historian, on what seem to be the good reasons for thinking something or the other about these topics. Of course, our sense of the good reasons for thinking something can inform our search for historical philosophers' reasons, but we do not end our search when we have found what we count as good reasons; we have, in addition, to determine the historical philosopher's attitude towards candidate reasons for thinking something.

Plato frustrates the historian's goals. One cause of this is the dialogue form in which Plato writes,<sup>1</sup> which leads to every philosophical view being put in the voice of some character or the other, none of them named 'Plato'.<sup>2</sup> As interpreters, we might suppose the main speaker in each dialogue (whether Socrates, Parmenides, Timaeus, or an Eleatic or Athenian Visitor) to be asserting or arguing for things Plato believed, but then Plato seems to change his mind uncommonly often for a careful thinker. For example, if Plato is Socrates in the *Protagoras*, *Meno*, and *Republic*, then, when writing the first he believes that virtue is not teachable and that virtue is knowledge; when writing the second he concludes that virtue is not knowledge because not teachable; when writing the third he describes how virtue can be taught. If Plato is Timaeus in the *Timaeus* and Socrates in the *Phaedrus*, then, when he writes the first he believes that the tripartition of the soul is the result of embodiment, but when he writes the second he believes that it is a cause of embodiment.

But after all, philosophers do change their minds, so we might accept that Plato held these various views at different stages and seek intellectual

reasons for his changes of mind. Here again, Plato's texts frustrate. We know so little about the order in which Plato composed his dialogues that it is difficult to determine the direction of the supposed changes in view. One objective technique for establishing chronology, stylometry, establishes that a few dialogues conform to late fourth-century fashions (e.g. avoiding hiatus and preferring '*kata*' instead of '*en*'). So we can identify some dialogues as written late in Plato's career by their stylistic similarities to the *Laws* (which Aristotle, *Politics* II.1264b26, says was a later work) and other late fourth-century works, and infer that other dialogues were written a little earlier or much earlier based on how much they differ stylistically from the late dialogues.<sup>3</sup> But a more fine-grained chronology is difficult. Plato sets his dialogues in the fifth century, so there are few references to contemporary events by which we might date them. Plato's *Menexenus* refers to the King's Peace of 386 BCE, and so must have been composed after that date, but this isolated bit of information does not enable us to establish a chronology when we do not know what Plato wrote before or after it. And, given the vastly different judgements of philosophical progress in Plato found in the history of scholarship, it does not seem that we should rely on our judgements of philosophical progress across the dialogues to suggest an order of composition.<sup>4</sup>

Many scholars now take the view that Plato deliberately left himself out of his writings in order to make his readers rely on reason, rather than his authority, to come to conclusions.<sup>5</sup> The right way to read the dialogues, on this view, is to participate in them philosophically—assessing positions and arguments ourselves.<sup>6</sup> But while 'following the argument wherever it leads' is necessary for every reading from the beginner's to the specialist's, it is as a historian seeking to understand what Plato thought that I go back to the dialogues—whole cloth, and not only the good bits—again and again. And to understand what Plato thought, it isn't enough—or really even possible—to engage in a private conversation with a dialogue; one must reconstruct the intellectual context in which Plato has his characters say what they say, including assumptions that we would not accept, such as that a cause cannot bring about an effect of an opposite nature to it. This is a different intellectual context from the one non-historian philosophers construct when they read Plato's dialogues to see what useful insights they might get from them on a particular contemporary topic or problem, a project for which

a rational reconstruction, using only grounds we can accept, would be appropriate.<sup>7</sup> But reconstructing the historical context can enrich the contemporary use of Plato's dialogues by deepening contemporary understandings of why Plato's characters say what they do, and for that we have to think about Plato's authorship even as we respect his attitude to authority.

An increasingly influential approach to the interpretation of Plato is to read the dialogues as operating on two levels: at one level is the dialogue between Socrates (or another main speaker) and his interlocutor(s), and at a second level above that is the dialogue between Plato and his audience.<sup>8</sup> For the sake of completeness, we should note that in some works there is a narrator reporting the conversation to another auditor (e.g. *Symposium*, *Theaetetus*), or Socrates himself narrates directly to the audience of Plato's text (e.g. *Republic*). Two-level interpretations have some advantages over interpretations that identify Plato's voice with the main speaker's. They can give a point to Plato's use of characterization and dramatic context: for example, Socrates says *p* because he is talking to so-and-so in such-and-such a context, and *p* is especially effective in these circumstances. And they can allow Plato to have some distance from the main speaker: for example, Plato doesn't believe *p*, but believes that such-and-such a person would do well to hear *p*, or that such-and-such a position involves a commitment to, or would be justified by, *p*, or that Socrates is making a mistake in saying *p*, either to such-and-such a person, or absolutely, and Plato is using the conversation to draw attention to the mistake.

While my interpretive approach in this book shares with two-level interpretations the assumption that it is not the main speaker alone, but the dialogue as a whole, that 'speaks for Plato', I do not rely on the *characters* of the interlocutors to understand why the main speaker says what he does—although of course the psychological views I attribute to Plato will need to be consistent with his psychological characterization of the interlocutors. I have four reasons for making little use of the characterization of the interlocutors in developing my account of Plato's moral psychology. First, relying on characterization to explain what is said shifts our attention away from Plato's intellectual motivations for the philosophical views advanced in the dialogues and in the direction of psychological speculation about the therapeutic needs of fictional characters. But inferences from what interlocutors say or do to global

character assessments are precarious. Does the fact that Meno can't wait to ask Socrates whether he thinks virtue is teachable (*Meno* 70a) show that he is tyrannical or philosophical? Does the fact that Adeimantus volunteers Glaucon as an example of the timocratic character (*Republic* 548d–549a) show that Glaucon is an honour-lover or does Socrates' reply that the timocrat is less musical than Glaucon show that the psychological theory is easily misapplied? Second, it would be question-begging for a work on Plato's moral psychology to understand Plato's characterization of interlocutors in terms of some one of his psychological theories, especially given the many caveats about the accuracy of these psychological theories Plato puts in the mouths of the main speakers (*Republic* 435c–d, 504a–b, 611e–612a; *Phaedrus* 246a, 265b–c; and, with regard to the whole account of the natural world, which includes souls, *Timaeus* 29c). Of course, a work's psychological theorizing ought to be consistent with the behaviour of its characters, but in fact the characters' behaviour is consistent with more than one theoretical position. Third, what is appropriate (or not) to say to a person of a certain type threatens to dissolve into the just-so: does the interlocutor need to be chastened or humoured, corrected or encouraged? Does the fact that some interlocutors are unmoved (Euthyphro) or irritated (Calicles, Thrasymachus) by Socrates show that they are deficient or that Socrates' method is deficient? Finally, appeal to characterization to explain what is said can be an aider and abettor of our restricting Plato's philosophical views to what lies within our own philosophical horizons: we may think certain philosophical positions are so obviously right or wrong that Plato must have seen this and so is only introducing them into discussion to help along some benighted soul, or to show that such a position is motivated by the interests typical of a person of a certain sort, but our first response, given our distance from Plato, ought to be to think ourselves into a perspective from which these claims appear true, and true for reasons. If we don't at least provisionally take seriously the philosophical positions defended in the dialogues, we risk getting out of the dialogues only what we already believed, along with an account of how Socrates or Parmenides or a clever visitor treated or should have treated some interlocutor who isn't as enlightened as we are.<sup>9</sup>

It may be objected that Plato himself directs us to understand Socrates' speeches in the dialogues in terms of their appropriateness to his interlocutor's character, for in the *Phaedrus* Socrates says that since speech is

meant to direct the soul, the expert in speech needs to know how to classify souls and determine for each kind of soul what kinds of speech are persuasive (271d), and to deceive skilfully, which requires departing from the truth as little as possible, which in turn requires knowledge of the truth and the ability to discriminate incremental departures from it (262a–b). If Socrates practises what he preaches (especially given the wide variety of characters among his interlocutors, from Callicles to Theaetetus, and given the above-noted diversity in the philosophical positions taken by Socrates), perhaps the dialogues represent different types of souls being given speeches suited to either teach or persuade them, as befits their character.

However, Socrates' self-representation across the dialogues and even in the *Phaedrus* itself suggests that he would be *unqualified* to practise what he preaches in the *Phaedrus*, because the expertise in speaking he describes there requires the kind of definitional knowledge he *lacks*. For example, in the *Phaedrus* he goes on to say:

Until one knows the truth about each thing concerning which one speaks or writes, and has become able to define each thing by itself, and knows, having defined it, how to divide it into kinds up to the point of further indivisibility, and in the same way concerning the nature of the soul, seeing thoroughly [and] discovering the form fitting to each soul, disposes and orders his speech thus, giving multi-coloured and multi-modal speeches to a multi-coloured soul and simple ones to a simple—not before this will one be able to handle the class of speeches expertly, insofar as its nature allows, either for teaching or for persuading, as our whole earlier discussion has shown. (277b5–c6; translations mine unless indicated otherwise; emphasis mine)

But Socrates has said at the outset that he does not even know whether he (i.e. his soul) is simple or complex (230a), and his exemplary speech about the goodness of love depends on a 'likely' account of the soul, it being available only to a god to know the truth about what sort of thing the soul is (246a).<sup>10</sup>

Of course, there is a weak sense in which Socrates says what's appropriate to his interlocutors: he finds conversation topics that his interlocutors have something to say about; he seems sensitive to what his interlocutor can understand and accept—for example, he is committed to answering questions in terms the interlocutor can understand (*Meno* 75c–d). But these are norms of discussion aimed at inquiry for anyone and everyone who engages in it; Socrates may be a model participant in discussions of

this sort, but he nowhere claims and indeed repeatedly disavows the knowledge of the truths that might entitle one to fit speeches to souls for the sake of teaching or persuading them.

This is not to say that Plato's rich characterization is just window dressing, as if it contributes nothing to the meaning of the *Hippias Minor* that Hippias is a know-it-all and to the meaning of the *Republic* that Thrasymachus is so angry. Quite the contrary: Plato's characterization of interlocutors enables him to put some position *p* of interest into the voice of a person who 'stands for' *p*, and by that device to identify reasons why *p* might appear true.<sup>11</sup> So, for example, although Hippias says that goodness is a matter of bravery and justice and so on, he himself, like most of us, pursues goodness by acquiring productive and performative skills that are useful or admired, which suggests that he believes, or is committed to believing, that goodness consists in the power to do what one wants when one wants. Again, Thrasymachus notes that what is considered just varies from one community to another but also suspects that in these various communities are powerful parties whose interests are served by at least some of what is locally considered just. Neither this thought, nor the outrage it provokes, are unfamiliar; locating both in Thrasymachus allows us to dispassionately examine it, its scope, the conception of the good that underlies the outrage, and the consequences of all these for how we ought to live. Plato models such a focus on dialectic by the example of Socrates, who takes up and examines views that are not his own in order to determine whether or not to accept a given *p*.

For these reasons, the interpretive approach I take in this book focuses on the dialectical dependence of what is said, especially by principal speakers, in order to determine Plato's relationship to various substantive philosophical theses *p* that are presented and argued for in the dialogues. (By 'dialectical' I mean 'pertaining to the particular argumentative context', not Plato's or Aristotle's 'testing definitions' or 'pertaining to the first principle and grounding the sciences', which I will indicate with a capital 'D'.) The basic idea is that the interpretation of sentences that seem to express psychological doctrine should be informed by an understanding of what role those sentences play in the dialogue: for or against what are they being used? Are they adopted because the interlocutor, or most people, or the main speaker himself, believes them? Or because they explain something that one of these parties believes? Because my discussion of Plato's psychology in subsequent chapters will use this

approach, I try in the remaining sections of this introductory chapter to win the reader's confidence in the approach by showing that if we pay attention to the dialectical dependence of Socrates' claims in the *Republic* defence of justice on Thrasymachus', Glaucon's, and Adeimantus' assumptions, two perennial interpretive puzzles are solved: first, how Book I is related to the rest of the book (Section 1.2); second, what it means that the city Socrates constructs is happy as a whole (Section 1.3).

These examples are intended to illustrate the general point that attention to dialectical dependence yields important results about what Plato thought about various *p*'s and why, and, even where it does not allow us to determine whether Plato believed some particular *p*, it still shows us many philosophically interesting things he thought *about p* and *p*'s non-obvious relations to other topics. One prominent kind of dialectical dependence in Plato's dialogues is found in cases in which Socrates or another main speaker examines a *p* that he does not know or believe, hypothesizing some plausible *q* that would explain *p*, and/or examining what follows from some *p*.<sup>12</sup> In such cases, the speaker's reason for concern with a particular *p* is often that intellectuals or most people maintain *p* or ought to, given the *q*, *r*, and *s* they maintain or act on.

One of Plato's great achievements as a philosopher is to have seen connections between the various *p*'s and *q*'s concerning topics that are not obviously connected. The dialectical dependence of the views arrived at in the dialogues need not be an obstacle to determining what Plato thought, but may be a resource. And in fact, when we start to pay attention to dialectical dependence, we find that Plato provides readers with very explicit signposts about the correct attitude to take to the views he discusses.

## 1.2 The Relationship between *Republic* I and II–X

In *Republic* I, Socrates solicits definitions of justice from his richly characterized and opinionated interlocutors and then refutes them. When Thrasymachus speaks in praise of injustice, Socrates rises to defend justice. But *Republic* I concludes in aporia, with Socrates saying that he does not know what justice is, and that not knowing what justice is, he is not in a position to know whether or not it is a virtue, or whether or not it makes its possessor happy (354a–c). In Books II–X, by contrast,

Socrates defines justice in the city and soul and then argues that the just person is happier than the unjust. Many students of Plato liken the Socrates of *Republic* I to the knowledge-disavowing and refutative Socrates of the 'early' dialogues (e.g. *Euthyphro*, *Lysis*) and contrast this with the Socrates of *Republic* II on, whom they see as dogmatic, or constructive; many also take Plato from *Republic* II on to be putting his own ideas into Socrates' mouth. In addition to these differences in Socrates' philosophical approach, scholars point to doctrinal innovations in *Republic* II–X: a tripartite psychology that recognizes irrational desires as an independent source of action and belief; an account of the non-rational education required to instill correct values into pre-rational children; a warning about the dangers of practising Dialectic without adequate preparation; replacement of the analogy between virtue and craft by an argument that justice is to be valued for its own sake and not only for its consequences.<sup>13</sup>

This characterization of the differences between *Republic* I and the rest already suggests an explanation: Plato uses Book I to point out Socrates' shortcomings and thereby to motivate his continuing in a different vein in *Republic* II–X.<sup>14</sup> Among the shortcomings, many scholars have identified the inadequacy of Socrates' elenchus, that is, his method of pointing out inconsistencies in the interlocutor's beliefs to motivate further investigation, for philosophical progress. Some scholars have found elenchus inadequate on the grounds that beliefs may persist even though they have been shown inconsistent, for example if they were acquired a-rationally (a possibility made vivid by the accounts of education in *Republic* II–III, of character degeneration in VIII–IX, and of poetry's effects on the soul in X) and/or if they are sustained by some irrational desire (as suggested by the tripartite psychology of IV and X).<sup>15</sup> A possible upshot is that if beliefs about justice are the product of an ideology that serves self-interested rulers (as Thrasymachus argues in *Republic* I), then elenctic investigation of our beliefs about justice is pointless.<sup>16</sup> In any case, elenctic discussion is limited by the interlocutors' (inferior) abilities.<sup>17</sup> For these and other reasons, agreement between Socrates and his interlocutors is no guarantee that the discussion is converging on the truth. So, it is argued, after *Republic* I Socrates no longer relies on elenchus but develops a new psychology, as well as a new metaphysics and epistemology, in the service of providing an account and justification of justice.



Yet it is dramatically incredible that the very same Socrates who fails to convince anyone at the end of *Republic* I because his philosophical method assumes that all his interlocutors' beliefs are changeable by rational argument should just moments later, without notice, launch into an account that makes entirely different assumptions about their beliefs.<sup>18</sup> How can (the character) Socrates suppose one moment that only rational factors determine what people believe and do, and the next moment suppose that people also believe and do what their irrational desires lead them to? Even if Socrates eventually lays out these new assumptions, this supposed change in method of argument would precede that exposition by two books.<sup>19</sup>

In fact, Plato himself provides a dramatically and philosophically superior connection between *Republic* I and the subsequent books than do the accounts just surveyed, beginning with Socrates' own diagnosis of the failure of the discussion in *Republic* I:

[Socrates, addressing Thrasymachus:] I did not feast well [viz. on our 'banquet of speeches'], thanks to myself, not you; but I seem to be like the greedy, who seize everything served to taste it before they have tasted the previous thing properly: before we investigated the first finding, what the just is, I left it behind, rushing off to investigate whether it is vice and foolishness or wisdom and virtue; when an argument came up that injustice is more profitable than justice, I didn't abstain from going into it [and] leaving the other one behind, with the result that I have come to know nothing (*mēden eidenai*) from the discussion. For if I don't know (*mē oida*) what justice is, I will scarcely know (*eisomai*) whether it'll turn out to be a virtue or not, and whether the one who has it is not happy or happy.

(354a13–c3)

Here, in saying that he shouldn't have argued that justice is a virtue or that its possessor is happy when he had not yet defined justice properly, because in the absence of an adequate account of what justice is, he cannot know what it is like, Socrates makes a methodological point familiar from many other dialogues, often made in criticism of his interlocutor: when Polus praises rhetoric, saying what it is like before saying what it is (*Gorgias* 448e, cf. *Meno* 71b, *Laches* 190b); when Meno gives instances of virtue instead of a definition of virtue, or when he defines virtue by its parts (*Meno* 72a–b, 73e–74a, 79d–e); and when Theaetetus gives examples of knowledge instead of a definition of knowledge (*Theaetetus* 146d, cf. *Euthyphro* 6d–e). Socrates is saying here in the *Republic* that one needs knowledge of what F is in order to know

whether F is G, H, I, and so on, and one needs knowledge of F in order to know whether a, b, or c are instances of F. This is not to say one is a cognitive blank about F prior to knowing what F is;<sup>20</sup> rather, it is a claim about the structure of knowledge of a subject and the conditions for making authoritative statements about it. Our interpretation should not ignore Socrates' own diagnosis of his own failure, that it is his lack of knowledge of justice that's to blame, even if we also think his grasp of human psychology—of why people believe what they do, stray from what they ought, rationally, to believe—inadequate. Socrates' self-diagnosis does implicate his claims that justice is (like) a craft and that justice makes its possessor happy. But this is not Plato criticizing Socrates over Socrates' head; it is Socrates' admission of a fault in his own procedure. A final reason to take Socrates' diagnosis as authoritative is that Plato has chosen to make Socrates not only the main character, but also the narrator of this dialogue.

But all this might seem to strengthen the case that the Socrates of *Republic* II–X breaks with the Socrates of *Republic* I.<sup>21</sup> How can the same Socrates be in aporia about justice one moment and then, just a few dramatic moments later, be able to give an account of what justice is, in the city and the soul, in order to say that it is a virtue and that its possessor is happier than the unjust person? Isn't what he does in *Republic* II–X in violation of the methodological principle he states at the end of *Republic* I?<sup>22</sup>

I will be arguing that it is not. To begin with, consider a parallel: in the *Meno*, although Socrates says he won't know the answer to Meno's question whether or not virtue is teachable until he knows what virtue is (71b, 86d), when Meno presses him with his question, he shows how one can investigate the teachability of virtue in the absence of knowledge of what virtue is: on the hypothesis (*ex hupotheseôs*, 86e3–4; *hupothemenoi*, 87b3–4) that virtue is knowledge, it turns out that virtue is indeed teachable, and on the hypothesis (*hupothesis*, 87d3) that virtue is what makes us good and so is always beneficial, if only knowledge is always beneficial, it turns out that virtue is knowledge. We have reason to suppose that only knowledge is always beneficial because we notice that without knowledge, things we ordinarily think are good are not beneficial; for example, courage without knowledge is rashness (86e–89a). Socrates says he models this hypothetical investigation of the teachability of virtue on the way a geometer investigates the question, 'can a given

area (x) be inscribed in the form of a triangle in a given circle (specified by its diameter (y))?' The answer is, 'yes, if that area (x) is the area of a rectangle that falls short of the rectangle constructed on the circle's diameter (y) by a rectangle similar to itself (i.e. such that the side shared by the two rectangles is their mean proportional)'.<sup>23</sup> We can see in both these cases that the hypothesis not only answers the initial question but also explains why the answer to the initial question is what it is. (So the literal sense of *hupotithêmi*, to place something under something else, may still be alive in Plato's use (cf. *Timaeus* 92a, which describes the god placing (*hupotithentos*) many supports under foolish animals so that they are drawn to the ground).)

The *Phaedo* gives another, fuller description of the use of hypothesis in the absence of definitional knowledge. Here, Socrates describes his method of investigation as follows: first, 'hypothesizing (*hupothemenos*) in each case the account (*logos*) that I judge strongest, I put down (*tithêmi*) as being true those things which seem to me to agree (*sumphônêin*) with it, concerning both cause and all other things, and as not true, those things which do not . . .' (100a) . . . Second, 'if someone should take hold of (*echoito*) the hypothesis itself, you would not let it go and would not answer until you could examine all the consequences from it (*ta ap' ekeinês hormêthenta*), if they seemed to you to agree (*sumphônêi*) with one another or to disagree (*diaphônêi*). Finally, ' . . . if indeed you should have to give an account of it, you would give it in this way, hypothesizing (*hupothemenos*) again another hypothesis (*hupothessin*) the one that appeared best of those above, until you should come to something sufficient (*hikanon*)' (101d). As Socrates illustrates the method in the *Phaedo*, to answer the question, 'why is this body hot?', he first takes up the most compelling answer as a hypothesis, namely, 'it participates in the hot-itself', which means taking it and other things that 'agree' with it to be (provisionally) true; for example, 'all hot things participate in the hot-itself'. Second, he examines the consequences, for example 'since the hot-itself excludes the cold-itself, the hotness in things must recede at the advance of coldness'; 'since it belongs to the nature of fire to participate in the hot-itself, no fire is cold', and so on, for consistency. If he finds them consistent, he (third) looks for an account of his hypothesis, such as, 'the hot-itself is . . .' and once again takes up the most compelling account, as a hypothesis, which means examining its consequences for consistency . . . and he keeps doing this until he comes to something 'sufficient'.<sup>24</sup>

These examples from the *Phaedo* and the *Meno* provide a precedent for explaining the *Republic* I–II change in how much Socrates says to positively characterize justice (in the soul and city) without attributing to him an entirely new persona, outlook about philosophical discussion, or psychological theory. How is it that in the *Meno*, one and the same Socrates on the one hand insists that one can't know anything about virtue, including what are its instances, unless one knows what virtue is, and on the other hand pursues the 'is virtue teachable?' question, apparently under similar practical pressure, without an answer to 'what is virtue?' In the *Meno*, we simply take him at his word that he is pursuing the teachability question on a hypothesis. Might he similarly be pursuing 'is justice better than injustice?' on a hypothesis in the *Republic*?<sup>25</sup>

Here are some reasons to think that the answer to this last question is 'yes'. The Socrates of Book II on repeatedly reminds his interlocutors that the results of his arguments are less than fully accurate (435c–d, 504b), and that he himself lacks knowledge, especially the knowledge of the good (506c) required for someone to be able to see whether anything (e.g. justice) is good (520c). So if he hasn't acquired knowledge, *how* is he able to answer the challenge renewed by Glaucon and Adeimantus in Book II to show that the just person is happier than the unjust person, not only for the sake of the consequences of being seen to be just, but because it is so good to be just?

We can see from the text of the *Republic* *what* he does.

- (1) He gives an account of what justice is in a soul, viz. the condition in which each part does its own work, and shows how this condition of the soul is good for its possessor (441e–442a, 443c–445b; Book V–VII's description of the philosopher; 586e).

*How* does he come up with (1), the account of what justice is in a soul, in the absence of knowledge of justice, in apparent violation of the priority of definition? (That there is a justice-itself, which would be what the definition of justice applies to, is attested at *Phaedrus* 247d.) Again, we can say *what* he does.

- (2) He supposes that justice in the soul is like justice in a city (368e, 435a–b), a supposition he supports by
  - (2a) an argument that the soul is like the city in the number and character of its parts (435b–441c), and

- (2b) an account of what justice is in the city: the condition in which each class does its own work, this being the work that it is best suited to do (432d–433e).

*How* does he come up with (2b), the account of what justice is in the city, in the absence of knowledge of what justice is?

- (3) He constructs a city that meets certain specifications:

- (3a) one in which the law aims at the good of the whole city (rather than of any one class, or the rulers, as Thrasymachus claims).
- (3ai) that (3a) should be the city they consider seems best to Socrates because it conforms to his functional account of the city: a city's function is to meet the needs of citizens (369b).<sup>26</sup>
- (3aii) Thrasymachus (338d & ff.) and Glaucon (359a), according to whom justice in a city is defined by the law, can accept that (3a) will yield an account of justice in the city.
- (3b) the benefits to citizens for the sake of which the city exists go beyond protection from being harmed (the only good Glaucon mentions, 358e–359b), and include material goods to fulfil bodily needs, as well as physical, musical, mathematical, and Dialectical education for the sake of bodily and psychological well-being.

From (3a) Socrates derives (2b), which, as he earlier said (368d–e), is easier to come by than (1).

Notice that Socrates doesn't give much of an argument for (2b) but says that it has been 'rolling around before our feet from the beginning' (432d, cf. 369b–370a)—although he does give some confirmatory arguments (justice is the 'remaining' virtue after moderation, wisdom, and courage have been found, the one that makes it possible for them to arise in the city and preserves their existence (433b–c); justice understood as doing one's own corresponds to the justice of judges, which consists in ensuring that citizens have their own (433e–434a)). Given (2b) and the assumption that justice in the soul has the same form as justice in the city (2, 2a), justice in the soul (1) will be each part of the soul doing its own work. And having a soul in which each part does its own work turns out to be a condition of harmony, which is the health of the soul, which

affords its possessor the enjoyments of the best goods (knowledge of Forms, the best pleasures of appetite and spirit) possible.

Each of these steps is open to criticism: one might say against (3), as Aristotle does, that the institutions of the city Socrates describes fail to make the city as a whole happy. One might say against (2) that justice is not the same in the city and soul, perhaps on the ground that contrary to (2a), the soul doesn't have the same parts as the city. Establishing (1) would require answering these and other criticisms.

The parallel between (1) being a consequence of the hypothesis (2), and the *Meno*'s 'virtue is teachable' being a consequence of the hypothesis 'virtue is knowledge', would be obvious but for three things. First, while in the *Meno* Socrates makes it explicit at the outset of his investigation as to whether virtue is teachable that what he is doing is to investigate a property of virtue in the absence of knowledge of what virtue is, in the *Republic* the setup of challenge and defence tends to overshadow the investigative nature of Socrates' discussion. Second, whereas in the *Meno* Socrates' hypothetical investigation is of whether an F is G in the absence of knowledge of what F is, in the *Republic* his investigation is of an instance of F-ness (the justice of the soul) in the absence of knowledge of what F (justice itself) is. Third, there is no explicit parallel in the *Meno*'s 'virtue is knowledge' hypothesis to the account of the city in the *Republic*, but it is the analogy between the soul and city that Socrates uses to come up with the account of justice in city and thus soul. (However, the analogy between soul and city does correspond to something in the *Meno*'s geometrical example, namely the similar rectangles, and it's reasonable to assume that in the geometrical case a diagram helps in finding the solution as does the sketch of the city in the *Republic*.) A later passage in the *Republic* enables us to see Socrates' reasoning about justice in *Republic* II–IX in terms of hypothetical investigation, and further, to locate the status of the city, the source of his account of justice, in this reasoning. In *Republic* VI, Socrates enumerates our cognitive powers and the objects to which they are directed by likening these to four subsections of a line:<sup>27</sup>

Now look again at how one should divide the division of the intelligible . . . : in which, on the one hand [alternatively, in one subdivision] the soul, using as likenesses the things then imitated [i.e. ordinary perceptible objects]<sup>28</sup> is compelled to inquire from hypotheses, not making its way to a principle but to a conclusion, and on the other hand [alternatively, in the other subdivision]—to

the unhypothetical principle—going from a hypothesis and without the images around it, making the investigation through forms using forms... I think you know that those who busy themselves with geometrical [problems] and calculations and other such things, hypothesizing the odd and even and shapes and three types of angles and other things akin to these appropriate to each investigation as if they know them, having made these hypotheses do not deem it worth giving an account concerning these either for themselves or for others, as if they were clear to everyone, and beginning from these and going through the rest they complete, consistently (*homologoumenōs*), that for which they initiated their inquiry... Then [I think you know] also that they use in addition visible forms and make statements about them, thinking not about them but about those things which these things are like, making statements for the sake of the square itself and the diagonal itself, but not [for the sake of] the one they draw, and likewise the other things they mould and draw, of which there are both shadows and likenesses in water, using these as likenesses, seeking to see those things which one cannot see otherwise than by thought (*dianoia*)... This, then, is the kind that I said was intelligible, but [I said that] the soul is compelled to use hypotheses for investigating about it, not going to a principle, as it is not able to go up higher from hypotheses, but using [as] likenesses those same things which have been imitated by the things below, in relation to which [things below] those [above] have been judged clear and valuable. (510b2–511a8)

In this presentation of hypothetical reasoning Socrates draws attention to the fact that mathematicians' use of hypothetical reasoning involves the use of diagrams. Socrates explains that although mathematicians make statements about the drawn square or triangle (e.g. 'let ABC be an isosceles triangle'), their reasoning is about/for the sake of the square itself or the circle itself.<sup>29</sup> Historian of mathematics Reviel Netz argues that in mathematical practice contemporary to Plato (which must be the background that makes Socrates' explanation above accessible to Glaucon, accounting for Socrates' use of active aorist participles to describe what the geometers do and Glaucon's ready recognition of what Socrates is talking about when he says what they do), a 'hypothesis' would have referred to the starting point of a mathematical proof, namely the act of 'drawing a diagram, declaring it to exemplify the case in question, and making the preliminary assumptions concerning it'.<sup>30</sup> While this does not fit the use of 'hypothesis' in the *Meno* or *Phaedo* exactly, it fits the Divided Line passage in the *Republic* like a glove. My proposal is that the city in the *Republic* functions as would a diagram in a mathematical hypothesis. The hypothesis is: 'let Kallipolis be a city whose law aims at the happiness of the whole'. It may be objected that the city can't function like a diagram in a geometrical proof

because it is not visible.<sup>31</sup> But Plato needn't think that what enables diagrams to play the role they do in proofs is that they are visible; he may think that it is their being *likenesses*, in respects which the verbal part of a proof points to. That is what the passage above, like the city-soul analogy, stresses. Further, geometers don't use ordinary perceptible objects as likenesses to help them think about the triangle itself or the square itself; rather, they construct schematically drawn triangles and squares to aid their thinking—and by saying that reflections and shadows are likenesses of these drawn objects (510e), Socrates is pointing out that the status of the drawn square is not that of a likeness of any perceptible square thing (such as a reflection or shadow one might see in nature) but is the same as that of the perceptible object itself. This is because both are likenesses of an abstract object to be investigated by thought.

Although the city in speech delivers an account of justice instantiated in a city that enables thinking about justice-itself, in the argument of the *Republic* Socrates uses it to think about another instance, justice in the soul. (It may be that such thinking is possible only via justice-itself, but justice-itself does not become an object of investigation in the dialogue.) To this extent, Socrates' procedure in the *Republic* is quite different from his investigation of the teachability of virtue in the *Meno*, where the hypothesis 'virtue is knowledge', since it is a proposed definition of virtue, does not stray far from definitional inquiry, and indeed its result directly contributes to definitional inquiry, for the result, 'if virtue is knowledge, then virtue is teachable' enables elimination of the candidate definition 'virtue is knowledge' because it turns out that virtue is not teachable. The hypothetical investigation of the *Republic*, by contrast, might suggest that justice is some sort of harmony of differentiated parts conducing to the good of the whole (since the just person doesn't differ from the just city with respect to the form of justice, 435a–b), but that suggestion is never tested. On the other hand, the consideration of one instance of something to answer a question about another does conform to the mathematical example of the *Meno*, where the question whether an area can be inscribed as a triangle in a circle of certain dimensions is answered by consideration of whether that area can be inscribed as a rectangle of certain dimensions.

I have been arguing that we can preserve the dramatic and philosophical unity of *Republic* I with the later books if we understand Socrates from *Republic* II on as answering the question, 'what is individual justice



and why is it good by itself for its possessor?' on a hypothesis, and I have proposed that this hypothesis is complex, consisting of a claim that the city is like the individual (stated explicitly by Socrates) and an account of the city the law of which aims at the happiness of the whole (assumed by Socrates on the basis of his discussion with Thrasymachus in Book I since Glaucon's challenge in Book II takes up Thrasymachus' position). Of course, I have not accounted for every apparent discontinuity and, in particular, I have not said anything about irrational desires. But this is because explaining what irrational desires are, how they are different from rational desires, and what the similarities and differences are between the irrational desires of the *Republic* and Socrates' assumptions about desires in the so-called 'early dialogues', involves substantive considerations about Plato's psychology, which I need to deal with in greater detail in subsequent chapters. Now, however, let us turn to a second interpretive puzzle illuminated by attention to dialectical dependence in the *Republic*.

### 1.3 Thrasymachus' Long Shadow: Justice, Law, and the Happiness of the Whole City

As Socrates concludes his description of the way of life of the guardians of the ideal city (= step (2) identified in Section 1.2),<sup>32</sup> Adeimantus objects that by depriving the guardians of personal property, Socrates isn't making them happy even though the city 'belongs to them':

And Adeimantus, taking up [the argument], said, Socrates, what will you say, if someone says to you that you are not making these men at all happy—and on account of themselves—to whom the city belongs in truth, but they enjoy no good of the city, like others who have acquired farms and build fine and great houses and acquire furnishings fitting for them, and make private sacrifices to gods, and entertain guests, and further, what you were now saying, possess gold and silver and all other things, however many are thought to belong to those who are to be blessed? Someone might say that the auxiliaries appear to be established in the city like mercenaries doing nothing but guarding it.

Yes, I said, and indeed they work for their food alone, receiving no wages in addition to food as do the others, with the result that if they wish to leave town privately, it isn't possible for them, nor [is it possible for them] to give [things] to mistresses, nor to use resources in any other way they wish, as those who are

thought to be happy do. You are leaving out of your accusation these and many other such things . . . I think that traveling the same path we will find out what we should say. For<sup>33</sup> we will say that it would be nothing amazing if these men are happiest just so; however, we are establishing the city not looking to this, how some one class will be surpassingly happy, but how the whole city will be, as far as possible. For we thought that most of all in this kind of city we would find justice, and injustice in the city that was established worst, and observing [these we would] discern what we were seeking long ago. Now, as we think, we are moulding a happy whole, not taking a few who are such and placing them in it. (419a1–420c4)

While scholarly attention to this passage has focused on answering the question of what it means for the city to be happy, and in particular what relationship this has to the happiness of its citizens (e.g. the city is a kind of super-organism whose happiness consists in its unity, and this happiness floats free of the happiness of its citizens;<sup>34</sup> a city's happiness is constituted by the happiness of its citizens<sup>35</sup>), this passage raises a prior question. When did Socrates and his interlocutors agree that their task was to establish a city looking to how the whole city would be as happy as possible? At first, Socrates was describing a 'city in theory' (369a); somehow, it comes to be 'the good and correct constitution' (449a). In particular, when did they agree that they would find justice most of all in *this* kind of city, and injustice in the city established worst?<sup>36</sup> Socrates' explicit reason for constructing a city was that a city, being larger than an individual person, would have more justice in it, which would make it easier to see (368e). Did that imply that the city Socrates was to construct would be the most just possible? Surely not: not knowing what justice in a city is, he isn't in a position to construct a maximally just city. More likely, he was supposed to construct the happiest city, on the assumption that it would be the most just. But why should his interlocutors agree that the happiest city is the most just, especially since they think that what is just is determined by the law, whether this serves the interest of the stronger, as Thrasymachus claims, or of the weak who agree not to do injustice in order not to suffer it, as Glaucon says (359a)?

Socrates says that a city comes to be because none of us is self-sufficient but all of us have various needs (369b) which it turns out are best met by a division of labour involving specialization and exchange, and assignment of labours according to ability (369e–370a). Consideration of what a city comes to be *for* suggests a non-conventional criterion

for determining goodness of a constitution; notice that Thrasymachus too, while a conventionalist about justice, accepts a (different) non-conventional criterion for the correctness of laws: whether or not they actually advance the interests of the rulers (339c, 340d–341a, cf. *Laws* IV 714c4).<sup>37</sup>

In Section 1.2 I proposed that since Thrasymachus says that the law, which defines justice, is in every actual city established to serve the interest of the rulers (338d), in Book II Socrates posits a city in which the law is established to serve the interest not of the rulers, but of the whole city, and asks, what is justice—that is, the lawful—in this city? To see how Plato is using Thrasymachus, it's as important to note what Socrates doesn't question in Thrasymachus' position as what he does. He never questions Thrasymachus' claim that the laws of existing constitutions are designed to benefit their rulers. Nor does he question that it is just to obey the law or that the law in a city defines what is just (whereas one might think the justice of obedience to the law depends on what the law is). Instead, he goes after Thrasymachus' claim that expertise in ruling enables its possessor to legislate for his own advantage—this, he argues, is not how it is with other expertises, which, being themselves non-deficient, seek the advantage of that object about which they are expert, as, for example, medicine seeks the advantage of the body (341c–342e). To Thrasymachus' counter-example—the shepherd fattens his sheep not for the sheep but for his own profit (343b)—he answers that since each expertise produces its own characteristic good, it must be a separate expertise, in wages, that benefits the shepherd and all the practitioners of the other expertises in the same way (346c–d; to conform to the rule that the expertises benefit their objects and not their practitioners, it would have to be someone other than the shepherd and other experts that possesses this expertise in wages, or the shepherd and other experts in their distinct capacity as experts in wages). Although Book I ends in *aporia* about what justice is, the argument that expertise aims at the good of its object is not aporetic, and so Socrates can assume in the remainder of the *Republic* that legislative expertise aims at the good of those legislated for.

In making the good of the whole city the aim of the law, Socrates does not at all require his interlocutors to give up their own conventionalism about justice (Thrasymachus and Glaucon both give conventionalist accounts of the origins of justice, although their accounts of whom the

conventions serve differ), but rather to consider what is implied by legislation's being an expertise. It may be helpful to put Socrates' strategy in Aristotle's terms. In the *Nicomachean Ethics*, Aristotle says that what is lawful is somehow just, for legislative expertise determines what things are lawful, and we call each of these just. Since the law aims at the common good of all or of the best or of those in power, we call those things which are productive and protective of the happiness of the political community 'just' (1129b12–19). We can think of Socrates as arguing on the model of all the expertises that the law should aim at the common good of all (the whole city, rather than the best or the powerful), and then calling 'just' those institutions that are productive of this common good.<sup>38</sup> Or even more minimally, we can think of Socrates as hypothesizing a law aimed at the common good, and investigating what justice would be, on this hypothetical law.

In setting up the city, Socrates' most significant departure from Thrasymachus is the distinction he draws between legislator and ruler. The founders and legislators for the city are Socrates and his interlocutors.<sup>39</sup> They determine the laws and patterns to which the poetry that educates the young must conform (380b–c, 383c); they legislate permissible conduct for a lover towards his beloved (403b–c) as well as permissible medical practice and judging (409d–410a); they determine the guardians' property allowances (417b), marriage institutions (459e–460a), conduct in warfare with other cities (471c), and so on. They also select as rulers those citizens who most care for the good of the city (412d–e) and they choose female guardians to live communally with the male guardians (458c–d). The rulers do not legislate, although they may issue orders (*epitattein*) in imitation of the laws (458c). The consequences of Socrates' distinction are enormous. Even if Socrates has discredited Thrasymachus' claim that the ruler qua ruler or ruler qua expert aims at his own advantage, Thrasymachus' and Glaucon's assumption of self-interested motivation implies that what rulers (or contracting parties in Glaucon's account) do and should do is legislate (or make agreements, in Glaucon's account) to their own advantage. But since Socrates et al. are doing the legislating, but they are not part of the city, they can make the aim of law the good of the whole city—or anything else for that matter—without running up against anyone's theories about human motivation.

The legislator–ruler distinction allows Socrates to enact a law compelling philosophers, who have been educated to know what is good and

fine, both by itself and in particular circumstances, to rule the ideal city (519c–520e). They are particularly well suited to rule not only because of this knowledge but also because they have better things to do than rule (521b). When Glaucon objects that to require them to rule is unjust to them, Socrates reminds him that the aim of the law is to make the city happy, and that the law does so by harmonizing the citizens with one another and persuading and compelling them to bestow on the city whatever good they can. The passage raises several issues, the most-discussed of which is that doing what justice requires should not detract from the philosopher-rulers' happiness, for that would undermine Socrates' whole defence of justice. So scholars argue that it does not make the philosophers less happy to rule, either because their bodily nature requires practical activity,<sup>40</sup> or because their love of the form of justice motivates them, either to imitate it where they can<sup>41</sup> or to express their understanding of justice in their actions.<sup>42</sup> But it also seems important that Socrates says they are both persuaded *and* compelled to rule, where the compulsion likely comes from the law.<sup>43</sup> While the idea that justice demands obedience to the law is widespread, its role in this passage is justified not by this widespread acceptance, but by the interlocutors' own view, from the beginning of the dialogue, that the law defines justice. And Glaucon pronounces it impossible for the philosophers to disobey the order to rule since they are just people being given a just order (520e). But in addition to the compulsion of law, Socrates imagines persuading the philosophers by saying to them the 'just things' that are the very considerations that recommended to the legislators that they enact a law compelling the philosophers to rule: that it is just to contribute to the common good what one does best, and they are indebted to the city for their education (520b–c).

## 1.4 Conclusion

What progress have we made with respect to the questions with which we began, about how to read Platonic dialogues to determine what Plato thought? I claimed that if we pay close attention to the dialectical role of various statements that *p* in the dialogues, we better appreciate Plato's attitude towards any *p* and in particular Plato's view of the relationship between any *p* and various *q*'s, *r*'s, and so on. I have argued that this approach to reading the *Republic* shows that Socrates' account of and

defence of justice is an answer based on a complex hypothesis: the soul and city are similar enough in structure that the law-prescribed condition which enables the happiness of a city identifies the condition in the soul that makes a soul happy. If this is right, then we should expect that in other work, Plato would explore this hypothesis in the way prescribed by the method of hypothesis: examining further consequences, and seeking higher hypotheses that could ground this one (e.g. what, in greater detail and accuracy, are the correct laws for a happy city? If the soul is by nature divided, are its elements like the classes in a city or like something else?). I believe that the *Laws* pursues the first of these questions, and the *Phaedrus* and *Timaeus* the second. In the following chapters, I turn to psychological statements in the dialogues, examining them in light of their dialectical dependence, and in some cases, as hypotheses.

## Notes

1. Because many followers of Socrates, not just Plato, wrote dialogues memorializing Socrates' way of life (called by Aristotle *Sokratikoi logoi*), and because the Socrateses that emerge from these dialogues are so diverse, it seems that rather than aiming to be historically accurate, the dialogues are, as A. Momigliano puts it, 'experiments . . . directed towards capturing the potentialities rather than the realities of individual lives' (*The Development of Greek Biography* (Cambridge, MA, 1993), 46). The essays in P. Vanderwaert, *The Socratic Movement* (Ithaca, NY, 1994) point to ancient authors who took this approach. In other words, a fictionalized conversation could reveal the type of person Socrates was (in the author's eyes) or Socrates' way of life (as the author understood it) by showing what he would say in various conversational circumstances. We may safely suppose that Plato used his Socratic conversations to depict the philosophical way of life as he conceived it (initially; later he would introduce other exemplary figures).
2. This makes it tempting to use (some of) the Platonic *Letters* to determine what Plato thought and why. Too often, however, the *Letters* are treated as providing authoritative biographical bases rather than philosophical arguments for Plato's opinions.
3. I am here following C. Kahn's cautious summary of the findings of the best stylometric research in his 'On Platonic Chronology', in J. Annas and C. Rowe (eds), *New Perspectives on Plato, Modern and Ancient* (Cambridge, MA, 2003), 93–127.

4. For example, writing in the early nineteenth century, Schleiermacher grouped the dialogues into three, beginning with the investigative dialogues (e.g. *Phaedrus*, *Protagoras*, *Parmenides*), continuing with dialogues treating knowledge (e.g. *Theaetetus*, *Sophist*, *Phaedo*, *Philebus*), and ending with the constructive dialogues (*Republic*, *Timaeus*, *Critias*). Arguing that this order is confirmed by the movement of thought from myth to science, Schleiermacher wagers, '... if well-grounded historical traces were to be found of an earlier composition of the *Republic* prior to any one of those preparatory dialogues, though none such has yet been found, and, what is more, *will not be found*, we could not avoid falling into the most serious contradiction with our judgement upon Plato, and we should be much embarrassed how to reconcile *this instance of unreason* with his vast intelligence' (*Philosophy of Plato and Aristotle*, tr. W. Dobson (New York, 1973), 44 emphasis mine). A century later, convinced of the philosophical superiority of the argumentatively sophisticated and metaphysically non-committal *Theaetetus* to the metaphysically burdened narrative of the *Timaeus*, G. E. L. Owen argued that the *Timaeus* was a middle-period dialogue, advancing views refuted by the *Theaetetus* ('The Place of the *Timaeus* in Plato's Dialogues', *Classical Quarterly*, 3 (1953), 79–95). However, H. Cherniss showed that the balance of the evidence was against this chronology ('The *Timaeus* in Relation to Plato's Later Dialogues', *American Journal of Philology*, 78 (1957), 225–66).
5. A. Laks usefully contrasts Plato's 'disegotisation' with the common Presocratic practice of inserting themselves into their work (by means of a *sphragis* [=seal], a personal address, an 'I', etc.), but argues that Plato's seal is unmistakable even without his name. Laks claims that by attributing to Socrates opinions that he did not maintain (psychic tripartition, the theory of the Forms), Plato encourages us to identify Socrates' voice with his own. But this is a provisional claim, for Laks ends by Foucauldizing Plato's anonymity: what is said in the dialogues is what is possible at a particular historical moment. ('"Qu'importe qui parle." Sur l'anonymat platonicien et ses antécédents', in C. Calame and R. Chartier (eds), *Identités d'auteur dans l'Antiquité et la tradition européenne* (Grenoble, 2004), 99–117.)
6. The idea that the dialogues model philosophical discussion and are meant to stimulate the reader or listener to participate in their philosophical activity (assessing arguments and philosophical positions, etc.) instead of determining what Plato thought is found in a number of essays in G. Press (ed.), *Who Speaks for Plato? Studies in Platonic Anonymity* (New York, 2000); D. Nails in particular ('Mouthpiece, schmouthpiece' says Plato wrote dialogues in order to encourage philosophers in the Academy to engage in philosophical discussion of the sort the dialogues depict, 25). In *Plato and the Play of Character* [= *Play of Character*], R. Blondell argues that putting views in the voices of

- characters who believe them or at least are willing to defend them encourages us as readers to become participants in the dialogue while at the same time allowing us enough distance to be critical (Cambridge, 2002), 39ff.
7. This distinction between rational reconstruction and history of philosophy draws on R. Rorty, 'The Historiography of Philosophy: Four Genres', in R. Rorty, J. Schneewind, and Q. Skinner (eds), *Philosophy in History* (Cambridge, 1995), 49–75.
  8. Recent examples of two-level readings are D. Sedley, *The Midwife of Platonism: Text and Subtext in Plato's Theaetetus* (Oxford, 2004); D. Scott, *Plato: Meno* (Cambridge, 2006) [= *Plato: Meno*]; R. Weiss, *Philosophers in the Republic: Plato's Two Paradigms* (Ithaca, NY, 2012).
  9. In George Eliot's *Impressions of Theophrastus Such* (ed. N. Henry, Iowa, 1994), Theophrastus remarks at the end of 'Looking Inward', in which he has made many an unkind observation about his countrymen, '... in noting the weaknesses of my acquaintances I am conscious of my fellowship with them. That a gratified sense of superiority is at the root of barbarous laughter may be at least half the truth. But there is a loving laughter in which the only recognised superiority is that of the ideal self, the God within, holding the mirror and the scourge for our own pettiness as well as our neighbours'. Socrates' comments on the young men who follow him because they enjoy witnessing elenctic take-downs or who imitate him because they enjoy performing them (*Apology* 23c), and his description of the pleasure of seeing one's friends refuted as one that is mixed with the pain of malice (*Philebus* 48b–50a), suggest that Plato seeks to promote Eliot's first kind of laughter no more than does Eliot's Theophrastus.
  10. One might think that in the *Republic* when Socrates answers Glaucon's request that he describe the power and forms of Dialectic by saying that although he is not lacking in zeal, Glaucon won't be able to follow him because he would be seeing the truth itself (533a), the implication is that Socrates knows the truth itself. But Socrates may only be saying that his own practice of Dialectic involves dealing with abstract objects forms (hypothesizing them, finding their consequences, examining their mutual consistency, and hypothesizing higher forms for them), whereas Glaucon cannot yet reason without the aid of images (characteristic of *dianoia*, 'thought', on which see Section 1.2, pp. 12–22). This weaker claim would be consistent with Socrates' earlier reluctance to state his opinion about the good on the grounds that he would be talking about something he didn't know (506c), and his reminder that his model of the sun and its relationship to the coming-to-be and visibility of natural things to indicate how intelligibles owe their being and being intelligible to the Good (506e–509b) consists of his opinions (*ta emoi dokounta*, 509c3).



11. In 'On Plato's *Politeia*', S. Menn claims that Glaucon and Adeimantus in the *Republic*, like the Spartan and Cretan in the *Laws*, stand for Plato's intended audience, namely people who base their political ideal on Sparta and Crete (*Proceedings of the Boston Area Colloquium in Ancient Philosophy*, 21 (2006), 1–55 at 24–6, 42–5). By contrast, I treat the interlocutors as standing not for real-world *types* but for tendencies and clusters of beliefs (which might co-exist with contrary tendencies and clusters) in real-world individuals.
12. R. Woolf argues that Socrates cannot cross-examine himself, because what he cross-examines is beliefs, but whereas believing involves taking to be true, examining a belief requires openness to truth or falsehood, so I cannot simultaneously both believe that *p* and be genuinely open to *p*'s being false ('Socratic Authority', in P. Remes and J. Sihvola (eds), *Ancient Philosophy of the Self* (Dordrecht, 2008), 77–107 at 86–92). As we'll see, however, Socrates' examination (and hence self-examination) isn't restricted to beliefs but includes hypotheses and their consequences, appearances, and so on.
13. Irwin, *Moral Theory*, 178–84; cf. Reeve, *Philosopher-Kings*, 19.
14. Irwin, *Plato's Ethics*, 169; Reeve, *Philosopher-Kings*, 23.
15. e.g. D. Scott 'Platonic Pessimism and Moral Education', *Oxford Studies in Ancient Philosophy*, XVII (1999), 15–36; cf. Blondell, *Play of Character*, 181, who takes the point to be that even skilful argument can be alienating and powerless against a commitment to injustice.
16. Reeve, *Philosopher-Kings*, 15. Pointless for the purposes of learning the truth, perhaps, but surely not for the purposes of ideology-critique.
17. Blondell, *Play of Character*, says that from Book II on, Socrates adopts the methods of the sophists (such as defending a view for the sake of argument), which are more cooperative and less 'impossible' for education than one-on-one elenchus (184–6), and which enable inquiry without anyone's ego being on the line (208; cf. Irwin, *Plato's Ethics*, 189).
18. Reeve recognizes and embraces what I have called dramatically incredible and explains: the Socrates of *Republic I* is a perverted philosopher, brought up in a bad city; the Socrates of *Republic II* on is the philosopher who would be brought up in the ideal city (*Philosopher-Kings*, 24).
19. C. Kahn, in 'Proleptic Composition in the *Republic*, Or Why *Republic I* was never a Separate Dialogue' (*Classical Quarterly*, 43 (1993), 131–42), highlights the many points in Book I that foreshadow the rest of the *Republic*: idealization of the ruler as possessor of ruling expertise (340e–341d); the claim that good rulers rule unwillingly (346e–347d); consideration of the consequences of justice and injustice in the afterlife (330d), and so on. But pointing out these very interesting connections doesn't address the apparent discontinuity. On the other hand, it does sufficiently rebut the suggestion that the (dubiously Platonic) *Clitophon* is an alternative beginning to the

*Republic* and as such is evidence of Book I's separateness from the rest. There are significant differences between these two supposed introductions: the *Clitophon* attributes the questions about justice's product to Clitophon rather than to Socrates (contrast *Clitophon* 409b–d and *Republic* 332c & ff.); it attributes the answer, 'friendship, that is, agreement, that is, shared knowledge' to a companion of Socrates (409d–e); this answer begs the question of the content of knowledge in a way similar to the *Euthydemus* (288e–292e) and a later passage in the *Republic* (505c); it attributes to Socrates the opinion that justice is helping friends and harming enemies (410b), which in Book I is Polemarchus', not Socrates' view. However, the main point of the *Clitophon*—Clitophon's complaint that while Socrates is good at inspiring people to pursue virtue, he leaves them high and dry when it comes to helping them figure out how to pursue it (presumably because he leaves them without a definition)—is addressed in the rest of the *Republic*. Still, the proleptic connections Kahn finds between *Republic* I and the rest are far tighter.

20. As P. T. Geach supposes in 'Plato's *Euthyphro*: Analysis and Commentary', *Monist*, 50 (1966), 369–82.
21. Socrates' assumption that justice is a virtue is also cited by Irwin, *Plato's Ethics*, 180, and Reeve, *Philosopher-Kings*, 22, as requiring a fresh start in Book II. I. Vasiliou writes that when Socrates says if he doesn't know what justice is, he doesn't know if it is a virtue, he only means that he doesn't know whether justice as it is ordinarily conceived is a virtue; see his *Aiming at Virtue in Plato* (Cambridge, 2008) [= *Aiming at Virtue*], 37–9. But this is ad hoc.
22. Vasiliou raises this question too, but his answer is that Socrates leaves the final determination of the content of virtue to the philosopher-rulers, as he does, e.g. for specifying the content of the musical education towards justice (*Aiming at Virtue*, 231–2). I agree with Vasiliou that the account of psychic justice at the end of Book IV is not an account of justice but only of what it is for an individual to be just; however, my reasons (see Section 1.2) are different.
23. My understanding of the geometrical example has been helped by S. Menn, 'Plato and the Method of Analysis', *Phronesis*, 47 (2002), 193–223. I take the initial hypothesis to be 'virtue is knowledge'. The analogue to 'yes, if x is the area of a rectangle . . .' is 'virtue is teachable if virtue is knowledge'. The way to establish results from hypothetical reasoning in geometry, according to Menn, is to reason from what is hypothesized either to a conclusion known to be false (thereby establishing the opposite of what was hypothesized) or to a conclusion known to be true and then, reversing direction, to reason from that known-to-be-true proposition to what was hypothesized. In the *Meno* Socrates establishes that virtue is not knowledge by showing that its

consequence, that virtue is teachable, is false. Had 'virtue is teachable' been known to be true, 'virtue is knowledge' might have been proved by reasoning from virtue's teachability to virtue's being knowledge.

24. In *Clitophon's Challenge: Dialectic in Plato's Meno, Phaedo, and Republic* (Oxford, 2015) [= *Clitophon's Challenge*], H. Benson initially characterizes the method of hypothesis as a mode of inquiry to be followed in the absence of someone who knows (finding whom is the hope motivating, but always dashed by, *elenchus*). The more detailed account of the method that emerges from Benson's detailed study is as follows:

In response to a question to which no-one present knows the answer,

- (1) Proof-stage 1: Identify a reduced question, the answer to which (= the hypothesis) will provide an answer to the original question.
- (2) Proof-stage 2: Show how the answer to the reduced question yields an answer to the original question.
- (3) Confirmation-stage 1: Test the consequences (*hormêthenta*) of the hypothesis for their agreement with each other.
- (4) Confirmation-stage 2: If the consequences in confirmation-stage (1) agree, identify a second reduced question, and show how the answer to the first reduced question can be obtained from the answer to this.

Repeat (2)–(4) for this second reduced question, and so on (151–2).

25. Of course one may instead claim a discontinuity in the Socrates of the *Meno* like the supposed discontinuity in the Socrates of the *Republic* against which I am arguing. Some proponents of such discontinuities reason that the Socrates of Plato's 'early' dialogues is the voice of the historical Socrates and at some point (in the 'middle' dialogues) becomes the voice of Plato. See e.g. Scott, *Plato: Meno*, and G. Vlastos, 'The Socratic Elenchus: Method is All', in M. Burnyeat (ed.), *Socratic Studies* (Cambridge, 1994) [= 'The Socratic Elenchus'], 1–37. But if this is what Plato is doing, he is doing it clumsily.
26. Thanks to John Proios for pressing this point.
27. According to G. Fine, 'Plato in the *Republic* places himself at L3 [viz. thought, *dianoia*]' ('Knowledge and Belief in *Republic* V–VII', in *Plato on Knowledge and Forms* (Oxford, 2003), 85–116 at 106). Fine identifies two criteria to distinguish thought and Dialectic: (i) the use of sensible images of forms (for the sake of thinking about the forms), and (ii) the movement from hypothesis to conclusion. In arguing that Socrates' defence of justice in the *Republic* is an instance of dianoetic reasoning, I am disagreeing with Benson's view that Plato regards *dianoia* as incorrect use of the hypothetical method (*Clitophon's Challenge*, 241–8). The dianoetic sciences (e.g. the mathematical sciences to be studied by future rulers of Kallipolis, *Republic* 533d) simply do not go all the way to a first principle, and they treat hypotheses as first principles—as one would expect of a subordinate science.

28. The mss. except A say *tmêtheisin*, but Adam argues for *mimeithesin* (which is also what Proclus reads) on the grounds that the former would include the objects of *eikasia* in what the mathematicians are using as images, but 510e says that shadows and reflections (i.e. the objects of *eikasia*) are imitations of the figures the mathematicians make and draw.
29. R. Robinson, *Plato's Earlier Dialectic* (Ithaca, NY, 1941), 154–6, argues that there is no connection between the use of hypotheses and diagrams: the geometers, he thinks, use visible images because they do not investigate hypothetically but dogmatically. But why should your use of images depend on whether you regard your starting points as true or hypothesized? Further, in our passage Socrates is *praising* geometers for the way they use diagrams. And the way they use diagrams—making claims about drawn things but for the sake of the square etc. itself—is the way Socrates has just used the image of the sun to reason about the good's relationship to our understanding and the intelligibles.
30. R. Netz, 'How Propositions Begin: Towards an Interpretation of *hupothesis* in Plato's Divided Line', *Hyperboreus*, 9 (2003), 295–317, at 309.
31. This objection is due to Clerk Shaw.
32. At 372e, Socrates says that he considers the city in which productive labour was divided for efficiency in production, including considerations of scale of production and citizens' different abilities for different productive tasks (369e–370a), to be the true and healthy city, but then he agrees to Glaucon's demand that the city enjoy luxuries, introduces a military for protection, and then introduces the education required to make this class a class of true guardians—towards the end of which education he comments that they have been purifying the only recently luxurious city (399e).
33. Grube-Reeve (in J. Cooper and D. Hutchinson (eds), *Plato: Complete Works* (Indianapolis, 1997) [= *Complete Works*]) and Griffith (in G. Ferrari and T. Griffith, *Plato: The Republic* (Cambridge, 2000)) omit the connecting *gar* in their translation; P. Shorey (*The Republic/Plato* (Cambridge, MA, 1937)) and C. Reeve (*Plato's Republic* (Indianapolis, 2004)) translate it.
34. K. Popper, *The Open Society and Its Enemies: The Spell of Plato* (Princeton, 1962), 169; cf. 79–81; B. Williams, 'The Analogy of City and Soul in Plato's *Republic*', in E. Lee, A. Mourelatos, and R. Rorty (eds), *Exegesis and Argument: Essays in Honor of Gregory Vlastos* (Leiden, 1973) [= 'The Analogy of City and Soul'], 197.
35. G. Vlastos, 'The Theory of Social Justice in the *Polis* in Plato's *Republic*', in H. North (ed.), *Interpretations of Plato: A Swarthmore Symposium* (Leiden, 1977), 15; cf. J. Annas, *An Introduction to Plato's Republic* (Oxford, 1981) [= *Plato's Republic*], 179. My own account of the relationship between city's and citizens' happiness, viz. the city is happy to the extent that its citizens are

- happy and its constitution is the cause of their happiness, is a variant of this account. See R. Kamtekar, 'Social Justice and Happiness in the *Republic*: Plato's Two Principles', *History of Political Thought*, 22 (2001), 189–220, at 205–6.
36. At 376d, Socrates says their inquiry aims at finding the origins of justice and injustice in a city; at 372e he agrees to describe the luxurious city on the grounds that in it one can see injustice as well as justice coming to be; at 371e Adeimantus guesses that justice and injustice are to be found in the (first, simple) city in the citizens' need of one another.
  37. The author of the *Constitution of Athens*, although he is no fan of democracy, does seem to admire the fact that Athenian laws preserve Athenian democracy.
  38. In 'Ethics and Politics in Plato's Defense of Justice', in Mark McPherran (ed.), *Plato's Republic: A Critical Guide* (Cambridge, 2010), 65–82, I argued that it would also beg the question of the *Republic's* ethical argument if citizens educated to be just were said to be happy at this stage; however, a different explanation is needed for Socrates' repetition of the aim of the law at 519d–520a. The issue here is not avoiding question-begging, but remembering what the law aims at (which is separate from the issue of the natural relationship between justice and happiness).
  39. J. Annas counts over forty references to the law/lawgiving concerning the ideally virtuous city in the *Republic* and remarks, '...if an aim is an important one for the achievement of the ideally virtuous city, Plato thinks that no argument is needed for it to be taken as an appropriate object for legislation. The lawgiving metaphor embodies this assumption' ('Virtue and Law in the *Republic*', in R. Patterson, V. Karasmanis, and A. Hermann (eds), *Presocratics and Plato: Festschrift at Delphi in Honor of Charles Kahn* (Las Vegas, 2012), 165–82 at 168 n. 7, 172). My account in Section 1.3 is meant to explain why no argument is needed for the claim that legislation aims at the whole city's happiness.
  40. Reeve, *Philosopher-Kings*, 201–3.
  41. R. Kraut, 'Return to the Cave: *Republic* 519–21', *Proceedings of the Boston Area Colloquium in Ancient Philosophy*, 7 (1991), 43–62.
  42. Irwin, *Plato's Ethics*, 298–317.
  43. E. Brown, 'Justice and Compulsion for Plato's Philosopher-rulers', *Ancient Philosophy*, 20 (2000), 1–17.

## 2

# Psychology for Sophists

### 2.1 Introduction

In the *Protagoras* Socrates makes or argues for a number of claims:

- (i) virtue is one
- (ii) virtue is knowledge
- (iii) the good is pleasure
- (iv) we always do what we believe is the best of the things we can do.

This chapter argues that in the *Protagoras*, these claims function as hypotheses that, if true, would explain how Protagoras can teach virtue and make students good. This introductory section sketches the movement of the dialogue as a whole, providing reasons along the way to take (i) and (ii) as hypotheses that explain how Protagoras' claims to teach virtue could be true. Section 2.2 argues that (iv) is a 'higher hypothesis' grounding (ii) [and a fortiori (i)], and that (iii) is a 'highest hypothesis' grounding (iv) [and a fortiori (ii) and (i)]. The aim of Section 2.2 is to show that in the *Protagoras* Socrates does not assume (iv) without argument, but instead introduces it as an alternative to the popular conception of the soul as pulled in different directions by competing passions. Finally, Section 2.3 compares the type of psychological explanation promised by (iv) with the type of psychological explanation available to the popular conception of the soul.

It is worth addressing two preliminary objections at the outset. First, one might suppose that since Socrates is refuting Protagoras in the *Protagoras*, it is unlikely that he would also help Protagoras by developing hypotheses that would show how his claim to teach virtue could be true.<sup>1</sup> However, there is no incompatibility between refuting someone by showing them that their thesis is inconsistent with their other beliefs, and offering them alternatives that would save some subset of thesis and

other beliefs—not just ‘you’ll have to eliminate one of *p* or *q*’, but also ‘*p* would be more plausible if *t*’.<sup>2</sup>

Second, one might object that in the *Protagoras* Socrates tells Protagoras that he wants to examine Protagoras and himself, not any statements granted by Protagoras saying, ‘if you wish’ or ‘if it seems so to you’ (331c), suggesting that his standard for participation in philosophical discussions is only to say what you sincerely believe.<sup>3</sup> So how could Socrates be developing hypotheses on behalf of Protagoras instead of saying what he himself sincerely believes—unless he has a double standard? However, Socrates’ standards for discussion can’t rule out discussion of a hypothesis just because the discussants don’t believe it, for a little later in the dialogue, examining the view that someone might be temperate while acting unjustly (which Protagoras disowns but attributes to the Many), Socrates instructs Protagoras to answer on behalf of the Many whether he is giving his own opinion or not, because Socrates is in the first instance interested in examining the argument—although Protagoras and he may end up also being tested in the process (333c). I suggest that Socrates’ requirement is that when testing a thesis *p* (e.g. ‘wisdom and temperance are independent virtues’) by examining its consistency or inconsistency with other views *q*, *r*, *s* (e.g. ‘for each thing there is only one opposite’, ‘temperance is the opposite of folly’, ‘wisdom is the opposite of folly’), the parties grant those *other* views *q*, *r*, *s* only if they seem true to them. This constraint on other views is important because the consistency of *p* with a *q* that the parties think is as likely to be false as true wouldn’t give them reasons for or against *p*.

Turning now to the overall progression of the *Protagoras*: the dialogue divides fairly naturally into three parts, which I’ll call ‘Virtue is teachable’ (309a–328d), ‘Virtue would be teachable if one, and one if knowledge’ (330c–334c, 349b–351a), and ‘Virtue is knowledge if we always do what we judge best’ (351b–360d). In ‘Virtue is teachable’, Socrates first asks what of value a student can expect to get from studying with Protagoras. Protagoras answers that he makes his students successful in political and domestic affairs (319a, 328b, cf. 357e). Socrates calls this ‘political expertise’ (319a4) and wonders whether (the Athenians think) this can be taught.<sup>4</sup> In response, Protagoras relates a myth according to which Zeus ordained justice and respect by nature for all of us (320d–323a), then reviews existing practices of teaching virtue—adults give children examples of fine/just/pious and foul/unjust/impious actions and punish

unwilling learners; schoolteachers make students internalize the works of good poets as well as the musical modes and rhythms; and legislators give the laws as models for living (323c–328a)—and finally asserts that his own role is to teach those with uncommon political potential how to be most powerful in word and deed, in private and in public (328a–d). The myth shows how it might be that we all have a natural aptitude for political virtue, and the example of existing practices shows that virtue is widely (thought to be) taught, so there should be no surprise that Protagoras teaches virtue. When Protagoras is done speaking, Socrates says he has been persuaded that virtue is teachable (328e).

But while Protagoras' brilliant speech has suggested how we might be suited to virtue so that traditional teachers, and standing on their shoulders Protagoras, might make us virtuous, it hasn't said what virtue is such that it can be taught, and in particular by the methods Protagoras has described. As we saw in Chapter 1, in the *Meno* Socrates says that prior to knowledge of whether virtue is teachable is knowledge of what virtue is, which suggests that since Protagoras claims expertise in teaching virtue, he ought to be able to answer Socrates' questions about the nature of virtue (in 'Virtue would be teachable if one, and one if knowledge'). But Protagoras attempts to maintain that the various virtues are distinct and function independently of each other, and Socrates is able to refute each of his attempts. These refutations show that Protagoras likely does not have knowledge of virtue, but they do more than that. For if the virtues are ('many') independent powers whose recommendations for action can conflict (perhaps as the sense-organs, parts of a face, may deliver conflicting perceptual reports), as Protagoras maintains (329e–330b), it is difficult to see how Protagoras (or anyone) can make his students good; indeed, it is difficult to see how anyone can be good, period. For example, if one is courageous and does what courage demands, mightn't one's action nevertheless be foolish? Even though Socrates' refutations individually only expose contradictions in Protagoras' claims, their collective effect is to push Protagoras to accept that the virtues are all one. But Protagoras should welcome this, for if the virtues are one, Protagoras' guarantee that he will make his students good is on firmer ground than if the virtues are many and capable of conflicting. Further, if the way that the virtues are united is by being identical with wisdom, as is suggested by Socrates' arguments identifying temperance and courage with wisdom, then studying Protagoras' particular areas of expertise (good deliberation,



speech-making, literary criticism) seems much more likely to make students virtuous than if, for example, courage also requires a strong bodily constitution (as Protagoras himself supposes, 326c).

Again, as we saw in Chapter 1, in the *Meno* Socrates argues that virtue is teachable on the hypothesis that virtue is knowledge. After a meta-conversation about how Socrates and Protagoras should conduct their discussion (334d–349a),<sup>5</sup> Socrates briefly returns to the relationship between the virtues (349b) to argue that courage is wisdom on the grounds that wisdom is the difference-maker between confident actions that are fine and confident actions that are mad (349e–350c). When Protagoras objects (349c–351a), sketching an account of courage and confidence in the soul that explains why he thinks courage must be distinct from wisdom—although courage entails confidence, confidence can come from knowledge, passion, or madness; further, courage comes from nature and good upbringing of souls<sup>6</sup> (351a)—he finally brings to the fore what stands in the way of his accepting that virtue is knowledge: a view about human motivation according to which what we do depends on the relative strengths of various independent forces in our souls (which his account of existing practices teaching virtue must have assumed, since it included corporal punishment for those unwilling to be persuaded by instruction, 325d, and required a strong bodily constitution for courage, 326c). But this view about human motivation is inconsistent with his claim to have and teach a *technê* that guarantees that his students will be good and successful.

In ‘Virtue is knowledge if we always do what we judge best’, after ascertaining that Protagoras is committed to knowledge being the strongest force in the soul, Socrates observes that the Many do not agree with this for the reason that (*hoti aition esti* (352d7)), as they think, we are sometimes overcome (*hêttomenous, kratoumenous*) by pleasure, or pain, or some other passion, even though we have knowledge. How can Protagoras believe that knowledge is the strongest force of the soul if he also believes, like the Many, that we are moved by independent forces of independently varying strengths—for wouldn’t the latter allow knowledge sometimes to be defeated by pleasure? This is why, when Protagoras objects to their bothering with the Many’s view that knowledge can be overcome by pleasure, Socrates answers that it will help them find out how it is (*exeurein . . . pôs echei*) with courage and the other parts of virtue (353b1–3). Socrates returns to this topic after

introducing the fourth of his theses, psychological eudaemonism: that we always do what we believe is best (358c–360d). I'll argue now that in the *Protagoras*, (iv) is made plausible by Socrates' refutation of the Many's account of how knowledge can be overcome by pleasure, rather than—as both the mainstream account and its challengers have supposed—being required for the refutation.

## 2.2 The Introduction of Psychological Eudaemonism in the *Protagoras*

It is easier to follow the dialectic of the third part of the *Protagoras*, 'Virtue is knowledge if we always do what we judge best', if we divide it too, this time into four stages. In Stage I, Socrates asks Protagoras whether he agrees that living pleasantly is good, and that things are good insofar as they are pleasant. Protagoras, after demurring that a pleasant life is good only if the pleasures enjoyed are fine, and that pleasures can be good, neutral, or bad, says he'll agree if Socrates can show that pleasure and the good are the same (351b–e). Socrates then shows an imagined Many that they believe that pleasure is the good (353d–354e), for in addition to believing that pleasure is good, they have no other criterion of goodness than pleasure (354b). The Many do not avow their ethical hedonism—indeed, they call some pleasures bad (351c)—but have to be argued into seeing that they are, in fact, whatever they might have thought about themselves, ethical hedonists.

The difficulty of further specifying the Many's particular variety of ethical hedonism—present aim, or long-run, modal, and so on—may be due to its being a view they have defaulted to, or to its being a recent discovery that it is their view, or to both. (I will not enter into the dispute about whether Socrates himself accepts the hedonism in Stage I here, but I will return to the relationship between the Many's default ethical hedonism and Protagoras' views in Section 2.2.3, when I discuss Stage IV.) In Stage II, Socrates refutes the Many's view that we are sometimes overcome by pleasure and so do not do what we know to be best even when we are able to do it (353c4–356a5).<sup>7</sup> In Stage III, Socrates addresses a protreptic speech to the Many, urging them to learn the art of measurement (356a5–357e8). Finally, in Stage IV, Socrates explains to Protagoras and the other sophists what the relationship is between courage and wisdom

in terms of an alternative theory of human motivation, psychological eudaemonism (358a5–360d9).

### 2.2.1 Stage II: *Refutation of the Many* (353c4–356a5)

The aim of this part of the discussion is to show the Many that their ethical outlook, hedonism, is inconsistent with their account of a certain experience as the defeat of the agent's knowledge that one thing is best by the greater pleasantness of another. It is controversial what Socrates' argument is, so let us begin by reviewing the text:

I say that if this is so [sc. the good and the bad are pleasure and pain], your account becomes absurd, when you say that often,<sup>8</sup> *although*<sup>9</sup> a person knows of bad things that they are bad, he nevertheless does them, when he is able not to do them, *because* he is dragged and assailed (*agomenos kai ekplèttomenos*) by pleasures, and again you say that a person, *although* he knows, won't (*ouk ethelei*) do good things on account of immediate pleasures, *because* he is overcome (*hèttômenos*) by them. That these things are absurd will be clear if we do not use many names at the same time, 'pleasant' and 'painful' and 'good' and 'bad', but since these showed themselves to be two, we call them by two names, first 'good' and 'bad', later 'pleasant' and 'painful'. Now having put it (*themenoi*) thus, let us say that a person, *although* he knows of bad things that they are bad, nevertheless does them. If someone asks us, 'Why?' we will say, 'Because he is overcome'. 'By what?' he will ask us, and it is no longer possible to say 'by pleasure'—for it has changed to another name, 'the good', instead of 'pleasure'—so let us answer him and say that he is overcome . . . 'By what?' he will say. 'By the good,' we will say, 'for God's sake!' If the questioner happens to be arrogant, he will laugh and say, 'The thing you say is absurd, if someone does bad things, *although* he knows they are bad, not because he is compelled to do them, but because he is overcome by good things. Then,' he will say, 'in your opinion, do the bad things prevail (*nikan*) when the good things are not worthy (*ouk axiôn*) or worthy?' 'Clearly,' we will say in reply, 'not worthy. For [otherwise] he would not be mistaken (*ou gar an exêmartanen*), the one we say is overcome by pleasure.' 'But in virtue of what,' he will perhaps say, 'are good things unworthy of bad things or bad things unworthy of good? Is it in virtue of anything other than whenever the one are larger, and the other smaller? Or the one more, and the other less?' We are not able to say anything other than this. 'Then it's clear,' he will say, 'that this "being overcome" you speak of is taking larger evils as the price for (*anti*)<sup>10</sup> lesser goods.' Then that's that.

Let us change the names again to 'the pleasant' and 'painful' in the case of these same things, and let us say that the person does, then we said 'bad things', but now let us say 'painful things', although he knows that they are painful, because he is overcome by the pleasant things, although it is clear that they are unworthy to prevail. And what else is unworthiness in pleasure in relation to pain, other

than excess and deficiency compared to the other? And these are being larger and smaller, and more and fewer, and more and less, than the other. (355a5–356a5, my translation)

Two observations about this passage are necessary before we turn to the difficulties with previous interpretations that motivate searching for an alternative. First, Socrates' argument targets the Many's *account* of an experience (*pathos*, 353a1), not the experience itself. The Many are theorizing about, not merely reporting on, an experience. For the statement, 'we sometimes do not do what we *know* (*gignôskô*) is best even when we are able to do it, *because* we are defeated or ruled by a passion like pleasure or pain' (352d–e, cf. 352b; emphases mine), involves both a normative judgement that the cognitive state we are in is knowledge (not just belief or assumption), and a causal claim that pleasure or pain is the cause of our acting contrary to this cognitive state. Second, Socrates' stated goal is to show the Many that their account is absurd (*geloion/a*, 355a5, b4, d1) as preparatory to showing (*apodeixai/apodeixeis*, 354e5–8) them that the correct account is ignorance (cf. 353a, 357c–d). Both Socrates' aim and his strategy suggest that he is providing a refutation of the Many's view, which requires him to show the Many an inconsistency in their beliefs about the phenomenon, which would ideally motivate them to seek the truth of the matter about which they have just been found ignorant. Socrates' arguments against Protagoras' claim that the virtues are distinct in Part I of the dialogue take this form, for example, when Socrates argues that Protagoras' view that wisdom and temperance are distinct conflicts with the conjunction of Protagoras' beliefs that each opposite has only one opposite and that wisdom and temperance are both opposed to foolishness (332a–333a).<sup>11</sup>

However, most scholarly accounts of the absurdity do not turn up anything *the Many* should find absurd—which is what a refutation calls for. For example:

- (a) '[I]t is surely absurd to say that pleasure could prevent one from doing the most pleasant things.'<sup>12</sup>

Why is this 'surely absurd'? Is it because the desire for pleasure could not prevent one from choosing the most pleasant thing? But that is just what is at issue: why couldn't my desire for immediate pleasure overcome my desire for greater long-term pleasure? Alternatively, is it that 'pleasure causes less pleasant choices or actions' violates the principle 'like causes

like’?<sup>13</sup> But this principle is not mentioned in the text, and since explanations, even by Presocratic physicists, frequently violate this principle (as Socrates complains in the *Phaedo* 97a–b, 98e–99a, 101a–b), Socrates would surely need to make this point explicitly to the Many (as for example he does when he uses the principle at *Republic* I 335c–d).

- (b) ‘[T]aking greater evils/pains in exchange for greater goods/pleasures’ (i.e. what the agent actually does) entails ignorance, but by stipulation, the agent has been said to know.’<sup>14</sup>

But Socrates has to refute the Many’s account *before* he is entitled to use the alternative account that what the agent actually does entails ignorance. As he says to the Many *after* they have been refuted and are asking him what this experience is, ‘If we had said then [that is, before the refutation was complete] that the experience was ignorance (*amathia*), you would have laughed at us. But now if you laugh at us, you will be laughing at yourselves’ (357d1–3).

- (c) ‘When A chooses y over x, A chooses y because A believes y is pleasanter than x overall . . . But this belief conflicts with the belief . . . that the action being chosen is worse than the alternative.’<sup>15</sup>

But why are the *Many* absurd if they attribute conflicting beliefs (y is more pleasant than x; y is worse, and so less pleasant than, x) to A? Doesn’t Socrates attribute conflicting beliefs to his interlocutors all the time—indeed, isn’t he attributing conflicting beliefs to the Many in this very refutation?

(c) makes explicit a widespread assumption about the Many’s account that I wish to call into question, which is that the Many believe that ‘anytime A chooses x over y, it is because A believes x is pleasanter than y’. In this case, the Many accept *psychological* hedonism, the view that people always do what they believe is most pleasant. My reason for doubting that Socrates assumes psychological hedonism is that while he argues at some length in Stage I that the Many accept *ethical* hedonism, the view that pleasure is the good, he never argues that the Many accept psychological hedonism, the view that human beings pursue the greater pleasure in all their actions. And given that he has to argue to make explicit to the Many their implicit ethical hedonism, why wouldn’t he need to argue them into seeing they accept psychological hedonism, even if only implicitly? For it is crucial to the success

of Socrates' refutation *as a refutation* that the Many accept the views he shows to be inconsistent.

Some scholars have supposed that it's because *Plato* accepts psychological eudaemonism that he attributes psychological hedonism to the Many after proclaiming them ethical hedonists. In his influential 'Plato's *Protagoras* and Explanations of Weakness',<sup>16</sup> Gerasimos Santas says that in the Many's claim that the good in exchange for which the bad was taken is not worthy of (*ouk axiôn*) prevailing (*nikan*), the 'value term' (*ouk axiôn*) reveals that Socrates takes 'overcome (*hêttômenos*) by good', which is ambiguous between (i) 'the good is *worthy* of conquering the bad' and (ii) the 'desire for good (pleasure) . . . is *stronger* than the fear of . . . the bad (pain)', in the sense 'is *worthier*' (= i), rather than in the more plausible sense 'is *stronger*' (= ii) (13–14). Socrates does this (says Santas) because Plato 'suppose[s] that there is some consistent correlation between the agent's evaluation or rankings of the alternative[s] before him and the strength of the conflicting desires that attach to these alternatives' (25).<sup>17</sup> So the absurdity in the Many's account is that "overcome by pleasure" . . . contradicts the very principle of psychological hedonism' (20). So, according to Santas, while Socrates himself doesn't endorse the principle of psychological hedonism, it's the hedonism to which he objects; his own view is 'a principle, which Socrates (and Plato) argues for in many early dialogues, and which can be assumed here without begging any of the questions at issue: that every man desires or seeks (pursues) to get good things and seeks or desires to avoid getting bad or evil things . . .' (22).

As stated, this principle is too weak to play the required role in Socrates' refutation; what Santas requires is the stronger psychological principle that *in every action* the agent chooses what he believes is the better of the alternatives before him. But if Socrates' refutation relies on this psychological principle on the grounds that Socrates has argued for it elsewhere (and for the record, I can find no arguments for it in 'many early dialogues'), he isn't showing the Many that their position is absurd, only that it disagrees with his own position. Further, the assumption of this principle makes the ethical hedonism for which Socrates has argued at such length redundant: if we always do what we believe to be best, we can never act contrary to our knowledge (or indeed belief) as to what's best, no matter what (we believe) the good is.<sup>18</sup> Yet Socrates insists that the refutation of the Many *depends* on their belief that pleasure is the good (354e–355a).<sup>19</sup>

Finally, it is difficult to find extra-textual evidence that psychological eudaemonism (much less psychological hedonism) had the kind of popular currency in Socrates' or Plato's time that would have explained his assuming that it was the Many's view. It is better attested as an intellectual's viewpoint, but even here, the evidence is not clear. In the *Republic*, Glaucon bases his challenge to the value of justice on the assumption that doing what is best for oneself may involve injustice, so that a person who refrained from injustice even if it brought him no bad consequences (e.g. loss of reputation) would be thought a fool (360d). This may express only rational eudaemonism, the view that the intelligent person always does what seems to him best for himself, rather than psychological eudaemonism, the view that everyone does.<sup>20</sup> But even if it does express psychological eudaemonism, Glaucon says the view is espoused by 'Thrasymachus and a myriad others' (358c7–8) which may mean he attributes the view to certain intellectuals.<sup>21</sup> In Gorgias' *Defense of Palamedes*, a model of forensic rhetoric, 'Palamedes' asserts that everyone does everything for one of two things, to acquire some benefit or to avoid some harm (19), so that if he is so crazy as to betray the Greeks in full knowledge of the destructive consequences for himself, he cannot be so clever as to have carried out the betrayal in the first place (25). Here, 'Palamedes' asks that the standard of what a 'moderately sensible' person would be likely to do, as opposed to what a madman might do, be applied to determine whether he is likely to have done what he is accused of doing. Is the fact that the speech is notionally addressed to a popular jury a reason to suppose that ordinary people share its conception of motivation? Socrates' claim that he could not be so crazy as to intentionally make his associates harmful to himself (*Apology* 25e, discussed further in Chapter 3.2) seems not to have won over the majority of his jurors.<sup>22</sup>

A far more common psychological outlook than psychological hedonism/eudaemonism, found in works of tragedy actually addressed to a popular audience, conceives of the soul as a site in which various forces of diverse origins and of differing strengths come into contact and combine with or conflict with one another. In the brief overview of the dialogue in Section 2.1 we saw that this is how Plato's Protagoras thinks of the soul, and it seems most natural for a reader to suppose that this is also how the *Protagoras*' Many think of the soul. A couple of examples will help identify the outlook. In Euripides' *Medea*, Medea, pulled between vengeful anger against Jason and love for her children (1042ff.), concludes:

I am no longer able to look upon you [children], but am defeated (*nikômai*) by evils. / And I know what evils I am about to do: / but anger is master of my reasoning (*kreissôn tôn emôn bouleumatôn*), / that anger which is the cause of the greatest evils for mortals. (1076–80)

And indeed, Medea's deliberations—the plans she is setting to kill her children and Jason's new family, and then to escape—are all in the service of her anger.<sup>23</sup> Here Medea's action is the product of the strongest passion in her, and that passion commands her reasoning. In Euripides' *Hippolytus*, Phaedra, after she has confessed her shameful love for Hippolytus, theorizes that we do bad things not because of the badness of our thoughts, for many of us think well (*ou kata gnômês phusin prassein kakion; esti gar to g' eu phronein polloisin*); rather, we know the best things (*ta chrêst' epistamestha kai gignôsomen*) but don't do them, some of us because of laziness (*agrias hupo*), and others because we put some other pleasure before the noble (*hêdonên prothentes anti tou kalou allên tin'*) (373–83).<sup>24</sup>

This background suggests a clear extra-dialogical counterpart to the position that I am suggesting is represented by the Many (and assumed by Protagoras) in the *Protagoras*: 'being overcome', where *x*'s overcoming *y* is understood as *x*'s winning out by being stronger than *y*, would seem to be a popular way of thinking about the interaction between motivations so as to give rise to an action. If the Many's account is given in terms of strength, a successful refutation by Socrates should also be given in terms of strength.

#### 2.2.1.1 THE STRENGTH ACCOUNT OF THE REFUTATION OF THE MANY

In fact we can reconstruct Socrates' refutation (= Stage II) in terms of strength. Prior to the refutation, the Many hold:

- (M1): Pleasure (alone) is good and pain (alone) bad (taken over from Stage I).

Where an agent *A* is faced with choosing between a good/pleasure *x* and another good/pleasure *y*, it may happen that:

- (M2): *x* is better than *y*.  
 (M3): *A* knows *x* is better than *y*.  
 (M4): *A* goes for *y* because *y*'s pleasantness overcomes *A*/*x*'s goodness.



Here, (M4) explains A's behaviour and (M2) is the basis for criticizing A's behaviour.

Socrates reasons from the Many's premises:

- (S1) By (M1), if *x* is better than *y*, then *x* is more pleasant than *y*; if *x* is more pleasant than *y*, *x* is better than *y*.
- (S2) By (M3), *x* is better than *y*.

(S2) follows from (M3) because knowledge is factive.

- (S3) By (S1) and (S2), *x* is more pleasant than *y*.

But (S3) conflicts with (M4), for according to (M4), *y* is more pleasant than *x*.<sup>25</sup>

On this reconstruction the absurdity in the Many's account arises out of the conflict between their claim that A knows a fact, namely that *x* is better than *y*, and their claim that what explains A's action is a contrary fact, namely that *y* is more pleasant/better than *x*. It does not matter that A's judgement is made in terms of goodness and the Many's explanation is given in terms of pleasantness since these are, in fact, the same thing. What the Many have been convicted of is saying that *x* is better than *y*, because A knows it, and that *y* is better than *x*, because *y* overcomes *x*; or equivalently, that *x* is more pleasant than *y*, because A knows it, and that *y* is more pleasant than *x*, because *y* overcomes *x*. If we imagine the goodness or pleasantnesses of *x* and *y* each as a force of a certain quantum in the soul, we can see that it is absurd to say that a smaller force overcomes the agent rather than the larger one—unless we add in some other conditions—but that is just what the Many say when they criticize the agent for taking the lesser pleasure, having explained her action as due to the greater force of the pleasure she actually took. The identification of the pleasant and the good is crucial to this argument, for if *x* and *y* were to differ other than quantitatively, the Many would not be committed to this absurdity.

Understanding the argument as concerned with the strength of the pull of two pleasures, or two goods, shows that Socrates is not guilty of substituting co-referring terms, 'pleasant' and 'good', into an intensional context, as some scholars have complained.<sup>26</sup> The context isn't intensional. Not only does Socrates understand the Many's talk of 'overcoming' or 'defeating' in terms of strength, the most natural way to understand them;

at the end of our passage he also remarks that worthiness and unworthiness are nothing but comparative excess and deficiency in pleasure. Rather than ignoring the 'is stronger than' sense of 'overcomes' so as to irrelevantly attribute to the Many the 'is worthier than' sense, he is arguing that a hedonist's correct evaluation of which pleasure is most worthy can be nothing other than a correct report of the pleasure's magnitude, but the pleasure's magnitude determines the strength of its pull on the agent.<sup>27</sup>

Of course, in real life the strength of a pleasure's pull doesn't depend only on its magnitude, and Socrates' refutation doesn't imply that people never choose the lesser pleasure, or that whatever a person chooses affords him the actually greater pleasure. Rather, Socrates' refutation draws attention to the poverty of the Many's explanation, 'overcome by pleasure', in terms of strength: if I am choosing between the (actually more pleasant) *x* and (actually less pleasant) *y*, how is it possible for *y*'s greater strength to make me take it, even though *y*'s magnitude is lesser? Mustn't something other than *y*'s greater magnitude account for the greater strength of *y*'s pull on me?

In the transition to Stage III, Socrates anticipates an objection from the Many that suggests how a smaller magnitude might exert a greater pull: 'But, Socrates, the present pleasure is very different from the pleasant and painful at a later time.' In other words, *y*'s greater pull is due to *y*'s greater proximity to me, not *y*'s greater (absolute) magnitude. Socrates' reply to this—'Not by anything other than pleasure and pain, for it's not possible [for them to differ] by anything else' (356a5–8)—may seem to change the subject from the effects of the pleasures on the agent to the facts about the magnitudes of the pleasures, and Socrates may seem now to be making the irrelevant assertion that whether the pleasures of or consequent upon *x* or *y* are present or future, they still differ only quantitatively.<sup>28</sup> But in fact Socrates' reply is exactly on target: yes, in bringing up proximity or distance, the objection adduces a factor in addition to magnitude that can account for *A*'s choice of the lesser pleasure, but this factor affects the appearance of the pleasure's magnitude, not the pleasure's magnitude itself. And once appearances are brought in to explain (and not only to criticize) choices, Socrates will be able to show that our mistaken choices are due to the influence on us of false appearances and so due to our ignorance. In Stage III, Socrates will draw an analogy with the role of proximity in visual judgements of size to bring this out.

### 2.2.2 Stage III: *Protreptic* (356b1–357e8)

Stage III will conclude with Socrates saying to the Many, ‘if you want to live well, pursue the art that measures pleasure and pain’. As we just saw, in response to the objection that the present pleasure is different from the distant one, Socrates says they can’t in fact differ except quantitatively. He continues:

But like a good [viz. skilled] person, having collected together the pleasures and the pains, both the near and the far, set them on a scale, [and] say which of the two are greater. For if you set pleasures against pleasures, the greater should always be taken (*lêpteia*); but if pains against pains, the lesser and the smaller; and if pleasures against pains, if the pains are exceeded by the pleasures, even if the nearby [are exceeded] by the further or if the further [are exceeded] by the nearer, that action should be done (*prakteon*) in which these [greater pleasures] are . . .  
(356a9–c1)

One might reasonably ask how what one *should* take or do is relevant, here, given that the issue was what *happens* when someone seems to choose contrary to her knowledge of what is best or most pleasant.<sup>29</sup> For this reason, and on the grounds that Socrates must somewhere show that no one knowingly chooses less pleasure when he can choose more, some commentators take this passage not to be saying what should happen, but what must happen as a matter of psychological necessity, even though Greek *-teos* endings are imperatival.<sup>30</sup>

In context, Socrates is now in a good position to introduce ‘should’ talk. Now that the Many have been shown not only that it is absurd to think that knowledge of what is good can be overcome by pleasure given that pleasure is the good, but also that the only way pleasures differ from each other in fact is quantitatively, they are ready to be persuaded to seek the knowledge that would enable them to choose the greater pleasure and lesser pain in every situation. So Socrates reasons with them:

- (P1) [As the Many already believe,] pleasures and pains are the goods and bads in life, which differ only quantitatively, so one should always take/do the greater goods/pleasures.
- (P2) If our well-being depended on doing and taking more in some domain where there was an art of measurement, we would recognize this art of measurement as our salvation, for it would make appearances powerless (356c–357a).

Socrates asks the Many, 'Do things of the same size appear (*phainetai*) by sight (*têi opsêi*) to you bigger when nearer and smaller when further?' (356c5), thereby proposing that appearances are the cause of visual judgements that things of the same size are larger when close up and smaller when far off. Similarly, he suggests, appearances are the cause of our choices of lesser over greater pleasures. This proposal picks up on the objection (from Stage II), 'the immediate pleasure is very much different from the more distant' (356a–b): while the immediate and the distant pleasures aren't different in actual size, they are different in apparent size, just like nearby and far-off objects of the same size appear larger and smaller to vision.<sup>31</sup>

Seeing that Stage III is concerned with how we should, rather than how we do, make choices shows that Socrates' comparison between vision and choice is apt rather than question-begging. Vision is a domain in which an art of measurement, optics, makes appearances powerless or unauthoritative (356c). For example, Euclid's *Optics* explains not only that objects of equal size at unequal distances appear unequal in size, with the nearer object appearing larger, but also that the apparent size of an object varies with the size of the angle formed between the eye, as the apex of a cone describing the path of our visual rays, and the object, as base. Having studied optics, we have a way of determining the actual size of something from its apparent size, and when this determination conflicts with its apparent size we are able to make a judgement contrary to appearances (Socrates does not call the result of calculation 'appearance' too). We continue to have the visual appearances we had before we started, but we no longer form our beliefs about size based on those appearances, at least when it matters; rather, we use measurement and our knowledge of optics. If Socrates were assuming that we always do what we believe is best or most pleasant, and using the visual analogy here to explain that we choose among pleasures and pains whichever one appears largest, he would be making illegitimate use of the visual analogy, since visual appearances don't determine our choices. But he is not *explaining* how we make choices; he is *recommending* how we should make choices, namely by providing ourselves with the knowledge that would secure correct judgement. Just as optics doesn't cause us to choose the larger object but only enables us to determine which the larger object is, the art of measuring pleasure doesn't make us choose the more pleasant object; it only tells us which the more pleasant object is.

- (P3) Since (by (M1) and (P1)) our well-being depends on choosing pleasure over pain and more pleasure over less pleasure, it would be secured by an art for measuring pleasure and pain (357a–b).

The structure of Socrates' reasoning from (P1)–(P3) parallels the famous protreptic in Plato's *Euthydemus* (278d–282d), where, beginning with the widespread desire to live well, Socrates shows the young Clinias that he should pursue wisdom, since wisdom alone guarantees the correct use of things that enables living well. Socrates' protreptic there is supposed to show the sophists Euthydemus and Dionysodorus how one should persuade a young man to pursue wisdom (278d); similarly, Socrates puts a coda on his speech to the Many by saying to Protagoras and the other sophists, 'This is how we would have answered the Many' (358a). Just as there is a schema for refutation, there is one for protreptic: show how wisdom and/or virtue is the best means to the goods already recognized and desired by the interlocutor.

Notice that in (P2), Socrates is restricting the claim about the role of measurement to domains in which our well-being is at stake. Presumably for choices in which our well-being is not at stake, guessing, estimating, or randomly selecting may do as well. Think of the different use one would make of weighing when packing for a backpacking trip as opposed to for car camping.

Socrates adds to the analogy with optics some suggestions about how an art of measuring pleasures would make appearances of pleasure powerless:

Wouldn't the one [i.e. the power of appearance] lead us astray and make us frequently change our minds about the same things and regret [our] actions and choices of large and small, and wouldn't the art of measurement make the appearance powerless (*akuron*), and having shown us the truth, cause the soul to have peace, abiding in the truth (*menousan epi tōi alēthēi*), and save our life? (356d3–e2)

There seem to be two ways in which knowing the art of measurement would keep us from changing our minds and from regretting our actions and choices, and would give us peace of mind: first, it would guarantee that our judgement about the magnitude of pleasures is true and will not be falsified by a pleasure turning out, when one is 'overcome' by it, to have been greater; second, it would enable us to hold on to our true judgement about the magnitude of pleasure in the face of appearances to

the contrary. This means that the corrective that the art of measurement provides is not restricted to what we call akratic action, in which the mistake lies in the action, which we might suppose is done in the heat of the moment, and not in the previous, cool judgement. For the power of appearance unchecked by knowledge of measurement could also occasion a mistake in the prior, cool judgement.

It might be objected that my interpretation robs Socrates' argument of its philosophical interest. The reason this stretch of the *Protagoras* is interesting to anybody is that it offers a unique theoretical account of apparently akratic action, action contrary to one's judgement of what's best, a familiar and puzzling phenomenon the existence of which is a serious obstacle to the development of virtue. On my reading, the account says that when someone acts contrary to their judgement of what's best because of pleasure, either the pleasure was only apparently larger, and the fact that they were taken in by this appearance is evidence of their ignorance, or, if it really was larger, then their judgement of what's best must have been wrong, which is evidence of their ignorance. But, one might object, in the first place, these are two entirely different phenomena. In the case where I act contrary to my judgement of what's best, I will be angry or frustrated with myself for acting on or having motivations contrary to my judgement; I will feel conflicted and perhaps ashamed if I act on these motivations; afterwards, I will blame myself.<sup>32</sup> In the case where I make a mistake about a pleasure in prospect and end up enjoying it more than I expected when I give in, I might kick myself for my mistake and regret my efforts to avoid the pleasure, but this is very different.

But why shouldn't a new psychological theory, of appearances empowered by the agent's ignorance, explain phenomena apart from apparently akratic action? And why should the theory uphold our different responses to apparently akratic actions and to what we recognized as mistakes of judgement prior to the theory? Why shouldn't our responses fall into line with the theory, so that now, recognizing apparent *akrasia* as due to the power of appearance, I kick myself for my mistake and try to get my measurements right so I don't repeat it?

It might also be objected that my reading does not say enough about how ignorance is responsible for akratic action, and how knowledge prevents it—and that this is because I have been holding psychological eudaemonism at arm's length. But I am not done. Socrates' refutation

and protreptic don't explain how ignorance or knowledge or for that matter judgement bear on action; it's Stage IV of his argument, addressed to the sophists, that finally does this.

Socrates concludes the protreptic:

- (P4) [The Many have agreed that] going wrong in the choice of pleasures and pains is due to ignorance (*amathia*) of the art of measuring pleasure and pain (357d–e).
- (P5) So [the Many should] study with the sophists (357e).

Let's take stock before turning to the last stage of Socrates' argument. Socrates has removed a powerful reason for resisting the idea that knowledge rules in the soul, namely that sometimes pleasure overcomes knowledge (352d).<sup>33</sup> With this reason gone, the Many are now open to alternative ways of thinking about how the elements in the soul relate to one another, and Socrates has introduced a psychology involving appearances corrigible by knowledge. We may now turn to how it works.

### 2.2.3 Stage IV: *Explanation* (358a5–360d9)

Having concluded his speech to the Many, Socrates now returns to Protagoras and the other sophists; he will, as promised, resume the conversation about the virtues, in particular courage and wisdom, that the speech to the Many was supposed to subserve. But before this, he elicits their agreement to a series of statements:

- (E1) If the pleasant is [the] good, no one does things knowing (*eidôs*) or believing (*oiomenos*) that there are better things that are possible that he can do (358b8–c2).
- (E2) 'Being weaker than oneself' (*to hêttô einai hautou*) is ignorance and 'being stronger than oneself' (*kreittô hautou*) is wisdom (358c3–4).
- (E3) Ignorance is having a false belief about the most important things (358c4–5).
- (E4) No one goes willingly (*hekôn*) towards bad things or towards things he thinks are bad; it is not in human nature to want (*ethelein*) to go towards things one thinks are bad instead of (*anti*) good (358c6–d4).<sup>34</sup>

Although (E1) may look like a conclusion from the discussion with the Many, it isn't. That discussion depended crucially on knowledge being factive, in the refutation, and ensuring true judgement, in the protreptic, but (E1) makes a claim about belief, and belief is not like that. Some commentators minimize the 'or believes' in (E1) on the grounds that since (E3) defines ignorance in terms of false belief, the alternatives are still knowledge and ignorance.<sup>35</sup> But this doesn't help: where is the argument that one cannot act contrary to ignorance *or* false belief? These commentators are assuming that all our actions are determined by our beliefs about good and bad or better and worse in the first place, which, as we saw in Section 2.2.2, Socrates cannot assume is shared by the Many, and which his argument so far has not needed. But must we, at least here, attribute to Socrates psychological eudaemonism, the view that we always do what we believe is best? And if here, why not earlier?

Long ago, Gregory Vlastos complained that when Socrates says 'no one does an action knowing or believing that he can do something better' (the consequent of our (E1)) and 'no one does wrong willingly' (our (E4)), he is making an empirical claim that he thinks he has established deductively.<sup>36</sup> Indeed, Socrates says it 'isn't *in human nature* . . . to want to go towards what one thinks are bad instead of good things' (358d1–2).<sup>37</sup> Worse, (E1) suggests that this empirical claim is supposed to follow from a claim about value, namely that pleasure is the good—but how can a claim about human nature follow from a claim about value?

Since, as we have seen, Socrates' argument has not been deductive so far, he needn't think he has established an empirical claim deductively either. Instead, in (E1) he is introducing a principle of human motivation, psychological eudaemonism, by removing a major objection to it (*akrasia*) and by showing the poverty of ordinary ideas about motivations as forces in the soul. It's important that his argument has relied on ethical hedonism—without that, his case for the absurdity of the Many's account of apparent *akrasia*, which motivates the consideration of an alternative, does not get off the ground.

Protagoras has all the reasons to embrace (E1) that he had to accept 'virtue is knowledge' and 'virtue is one': for this is how the difference between success and failure, virtue and vice, can be entirely a matter of whether one's judgements are informed by knowledge or ignorance, entirely a matter of teaching. By removing the need to accept that there are motivations that compete with the agent's judgement about



goodness, Socrates has shown Protagoras how he can adopt a psychology that would support his professional claims.

### 2.2.3.1 IMPLICATIONS OF PSYCHOLOGICAL EUDAEMONISM'S HYPOTHETICAL STATUS IN THE *PROTAGORAS*

If psychological eudaemonism functions as a hypothesis argued to in the *Protagoras*, rather than as an assumption of Socrates' theory informing all his psychological statements in the dialogue, then perhaps the text does not offer any answers to certain philosophical questions that we might want to ask. Earlier (n. 7), I suggested that this was the case for the precise kind of hedonism Socrates attributes to the Many in the refutation. Another case, I believe, is questions about whether non-rational motivations exist or exercise motivational force, and if so, whether directly or indirectly, via their effect on beliefs. In *Socratic Moral Psychology*, Brickhouse and Smith argue that Socrates believes not simply that (as per the standard account of Socratic intellectualism) all motivating desires are dependent on beliefs about good and bad, but rather that all motivating desires are *directly* dependent on beliefs about good and bad, but may depend *indirectly* on non-rational desires, which are desires that do not update in response to our beliefs about good and bad, and can, by representing their objects as good or bad, influence our beliefs about good and bad.<sup>38</sup> These non-rational desires would have to be distinct from the passions, which as we saw are just beliefs about good and bad.<sup>39</sup> So are they appearances? For the *Protagoras*, Brickhouse and Smith say, 'the appearance of goodness is a feature of the sorts of things we desire . . . as a *result* of our desire'.<sup>40</sup> 'Something acquires the power of appearance when it becomes the object of a nonrational desire and so becomes recognized by an agent as a way to satisfy some appetite or passion—for example, as a pleasure or as a relief from some pain.'<sup>41</sup> On this account of Socrates' psychological theory, non-rational desires cause appearances, which cause beliefs about good and bad, which give rise to motivating desires. Now this can't be true generally, for Socrates thinks we have appearances where no desires are involved, for example about the relative sizes of two bodies at different distances from us. But in any case, because how appearances come about isn't relevant to the role played by the notion of appearance in Socrates' argument, it's quite likely that the *Protagoras* provides no determinate answer to the question, 'can non-rational desires influence beliefs and so motivate actions indirectly?'

## 2.3 A Comparison of Explanations on the Force and Appearance Models

So far I've argued that in the *Protagoras* Socrates makes room for psychological eudaemonism, the thesis (iv) that people always act in accordance with their judgement of what is best, by identifying pleasure with the good and thereby removing the most obvious candidate motivation to acting contrary to one's judgements of goodness. But once introduced, psychological eudaemonism can be seen to offer an intuitive and attractive model of psychological explanation. Plato demonstrates its explanatory appeal by a contrast, which we see when we step back from and reflect on the dialogue.

An experience familiar to readers of the *Protagoras* is of feeling slightly bamboozled by Socrates' first arguments against Protagoras' quite intuitive view that the virtues are independent and capable of conflicting. Like Zeno's arguments against plurality (in indirect defence of Parmenidean unity), they seem to be composed in the competitive spirit of showing that even greater absurdities follow from the intuitive thesis than the supposedly absurd one (cf. *Parmenides* 128c–d). Socrates' arguments that unacceptable consequences follow from the separateness of the virtues may undermine Protagoras, but they fail to persuade us. Readers suspect they contain tricks or fallacies; the arguments seem strangely formal; the general response they provoke is that however many problems with the separateness of the virtues these arguments turn up, our experience of people who seem, for example, courageous but self-indulgent or foolish trumps Socrates' arguments. Take, for example, his argument for the identity of moderation and wisdom (332a–333a):

- (F1) There is such a thing as folly.
- (F2) Wisdom and folly are opposites.
- (F3) When people act correctly and beneficially, they act moderately.
- (F4) People act moderately by moderation.
- (F5) People who do not act correctly act foolishly.
- (F6) People who do not act correctly do not act moderately.
- (F7) The opposite of acting foolishly is acting moderately.
- (F8) People act foolishly by folly.  
Support: just like (F4) people act moderately by moderation.
- (F9) When someone acts with some power *p*, she acts *p*-ly.

Support: if she acts with strength, for example, she acts strongly (induction).

- (F10) When someone acts *p*-ly, she acts from *p*.
- (F11) Each member of an opposition has only one opposite.  
Support: for example, ugly-beautiful, good-bad, and so on (induction).
- (F12) Something done oppositely is done from the opposite (F10).
- (F13) If people act foolishly by folly (F8), and acting foolishly is the opposite of acting moderately (F7), and one acts 'oppositely' from the opposite (F12), folly is the opposite of moderation.
- (F14) But (F13) conflicts with the conjunction of (F2) and (F11)—unless wisdom is the same as moderation.

In 'Dialectic and Virtue in Plato's *Protagoras*', James Allen helpfully describes what is wrong with these first arguments against the separateness of the virtues as the 'defect . . . [of] a failure to explain or illuminate'.<sup>42</sup> Allen contrasts this defect with the dialogue's discussion of action contrary to knowledge of what is best, which 'provides . . . at least the beginnings of an understanding of virtue, virtuous action and human good and their systematic relations . . . enough to begin to see what an argument about the relations among the virtues put together in the light of an understanding of virtue and the good might look like' (p. 28). For example, after introducing psychological eudaemonism (E1)–(E4), Socrates argues that courage and wisdom are identical:

- (E5) Fear is the expectation of evil (358d).
- (E6) By (E4) and (E5), no one goes willingly towards what he fears (358e).
- (E7) Neither cowards nor courageous people go towards things they think are fearful; both cowards and courageous people go towards things of which they are confident (359c–e).
- (E8) The actions of courageous people are fine and good; the actions of cowards shameful and bad (359e–360a).
- (E9) The fear and confidence of courageous people are fine and good; the fear and confidence of cowards shameful and bad (360a–b).
- (E10) The fear and confidence of cowards are shameful and bad because of their ignorance and error (360b).
- (E11) What makes a person a coward is cowardice (360c).
- (E12) Cowardice is ignorance about what is to be feared (360c).

- (E13) Cowardice is the opposite of courage (360d).
- (E14) Knowledge of what is to be feared is the opposite of ignorance about what is to be feared (360d).
- (E15) So knowledge of what is to be feared is courage (360d).

While Allen is right about the difference in explanation and illumination between arguments of these types, I think the cause of the difference is not quite what he says it is:

[T]he earlier arguments are dialectical in the Aristotelian sense because they rely on principles of the utmost abstractness and generality without taking into account the special features of the matter under discussion, namely human nature and action . . . The later arguments differ by coming to grips with the subject under discussion. When the general principles employed in the earlier arguments reappear, it is in the context of a conception, however partial and tentative, of human character and conduct that makes it possible to understand the conclusions reached with their aid and why they should be true *if* the conception of human nature and action on which they depend is correct. (30–1)

(Allen stresses ‘if’ because he thinks hedonism is only provisionally granted; I would extend this ‘if’ to the psychological eudaemonist thesis that we always act on our judgements of what’s best.)

But the difference between non-explanatory and explanatory Allen identifies doesn’t seem to be due to a move from general and abstract to concrete and subject-specific—nor should we expect Plato to share the view that more general and abstract principles would for that reason be less explanatory.<sup>43</sup> Rather, it seems to be due to the fact that explanations of actions and passions in terms of judgements of good and bad enable us to adopt the agent’s perspective. What (E5)–(E15) does, which none of Socrates’ earlier arguments against the distinctness of the virtues in the second part of the dialogue did, is to make the perspective of the coward or the courageous person relevant to the explanation of his actions and passions. Why do cowards break the line of defence in battle and run? Not just ‘by cowardice’ (as in the second part’s refutations), not because fear overwhelms them or their judgement (as on the ‘force’ model), but, if they are running from fear, because by (E5) they judge that not running will result in injury and death, which they judge to be great evils (greater than, for example, disgrace). And what could make it easier to understand the coward’s running than that he is trying to avoid what he believes to be (a greater) evil?

By contrast, the argument (F1)–(F14)—like all the arguments prior to the introduction of psychological eudaemonism—treats moderation, wisdom, and foolishness like the Many’s black-box forces, the agency of which is manifested through action-instances. Like heat, they give something of their own nature to their effects. Socrates’ examples of ‘that in us which’ leads us to act attest to this: if I took very little time to do something, is there a difference between describing it as ‘done quickly’ and ‘explaining’ it as ‘done by the agency of quickness’? Of course there’s no denying that if I took relatively little time, then I must be quick, but how much does ‘quickness’ explain? More to the point, if I did something by the agency of moderation, even after you learn this you might still want to know my reasons, or how I saw the situation, if you wanted to understand my action. Such understanding is promised by the principle of psychological eudaemonism introduced in the last part of the dialogue, for it allows us to explain my choice of this action that I judge this pleasure greater than that one, given that I judge that pleasure is the good.

While feeling pulled in different directions may be the norm for tragedy, it certainly isn’t the norm for life. When we are thinking about what to do, we routinely represent our choices to ourselves in terms of the relative goodness of our different alternatives, and try to understand the actions of others by trying to see how their alternatives look to them as they decide what to do. And there certainly seems to be a very robust connection between what we think is best, decide to do, and actually do. But it does also seem that not everything we do is the result of such decision-making, and so the question arises whether psychological eudaemonism is too strong a principle. But if psychological eudaemonism is only applied to our intelligent or rational actions, could we explain our foolish or irrational actions on the force model? We will return to these issues periodically in the rest of this book; for the present, however, let’s turn to another element in the Socratic intellectualist package, the unwillingness of wrongdoing.

## Appendix: Psychological Eudaemonism in the *Euthydemus*?

In the *Euthydemus* Socrates begins a display of protreptic speech by asking the young Clinias, ‘Do we human beings want (*boulometha*) to do well? Or is the question one of the ridiculous ones? . . . Who among human beings doesn’t want to do well?’ (278e3–6). That everyone wants to do well is a weaker view than

psychological eudaemonism, according to which doing well is the *only* ultimate end. Further, that everyone wants to do well is not the same as everyone believing they want to do well, much less doing all they do in the belief that they will thereby do well; that they want to do well may be something we conclude by observing their behaviour rather than soliciting their opinions. After obtaining Clinias' agreement that we do well by possessing good things, and enumerating the good things, Socrates corrects himself to say that we do well by possessing *and correctly using* good things, argues that both require wisdom, and concludes by telling Clinias he should seek this wisdom (278e–282d). (Later, though, they end up in aporia trying to identify what would be the content of this wisdom, and what good it would bring about, given their earlier agreement that wisdom must be the good-maker and the only good by itself, since other possessions are beneficial only when used with wisdom (288d–292e).)

In the course of the protreptic, Socrates says that wisdom makes human beings fortunate in every case; if one has wisdom, one has no need of good fortune in addition; if one has many good things, one is happy and fares well (280a–b). It may seem that wisdom can guarantee living well only on the assumption that we always do what we judge best (as in Socrates' psychological eudaemonist proposal at the end of the *Protagoras*). But Socrates concludes the argument with only the claim that wisdom is necessary for doing well and the source of rightness and good fortune (282a), and his distinction between possession and correct use of goods raises the issue of whether wisdom possessed is a guarantee of its own correct use. Certainly the argument is convincing that there is nothing better to pursue than wisdom, that wisdom can mitigate against bad fortune, and that not even good fortune in acquisition of goods guarantees their correct use. But this weaker conclusion doesn't require psychological eudaemonism.<sup>44</sup>

What can we learn from the dialectical context of this passage? First, Socrates' protreptic display is made to show the sophists Dionysodorus and Euthydemus, who claim to teach virtue, and to do it more quickly than anyone else (273d), how to recruit young men to be students. Their method is to refute a potential recruit when he answers a question by saying *p* as well as when he answers it by saying *not-p* (275b–277c). Socrates worries that their approach may cause the potential recruit Clinias' courage to fail (277d), explains that their refutations have exploited his ignorance in the correct use of words, and ventures that they have been making fun of him in a kind of initiatory rite (277d–e, 278c). This set-up is very similar to the *Protagoras*': Socrates is showing sophists who claim to make men virtuous by means of a verbal skill what are the presuppositions of their teaching and how they should approach potential students. That does not rule out his sharing these presuppositions, but he does not have their reasons to accept the presuppositions, since he does not claim to make men virtuous by teaching them a verbal skill.

## Notes

1. This objection was suggested by a question raised by Vanessa de Harven.
2. Indeed, this is also what Socrates does on Protagoras' behalf in the *Theaetetus*, when he introduces the Secret Doctrine to show how 'perception is knowledge' could be true. Thanks to Robbie Howton for pointing out this parallel.
3. Vlastos, 'The Socratic Elenchus', 1–37, at 2–3.
4. Socrates says: the Athenian democracy treats no one as an expert in political matters, and wise and good parents fail to make their children virtuous despite their interest in doing so (319b–320b). Why does Socrates treat the practice of the Athenians as evidence of the unteachability of virtue rather than of the Athenians' opinion that virtue is unteachable? He calls the Athenians 'wise', but is that because Protagoras' relativism commits Protagoras to the identity of what seems and is for the Athenians? Protagoras' relativism is only briefly in view in the *Protagoras* (334a–c), and in the *Theaetetus*, Socrates objects that Protagoras' identification of seeming and being does away with wisdom, in response to which Protagoras says that the wise person is the one who makes what is good seem and be for others (166d–167a). Alternatively, this may be the first of several moments in the *Protagoras* showing the dependence of the sophists' substantive positions about matters of value on the beliefs of the many. I discuss this dependence in 'Plato on Education and Art', in Gail Fine (ed.), *The Oxford Handbook to Plato* (Oxford, 2008), 336–59, and R. Kamtekar, 'The Profession of Friendship: Callicles, Democratic Politics and Rhetorical Education in Plato's *Gorgias*', *Ancient Philosophy*, 25 (2005) [= 'The Profession of Friendship'], 319–39.
5. Long speeches? Short question-and-answers? According to the discretion of the instructor, Protagoras? Or according to the capability of the inquirer, Socrates? As poetry is a traditional source for teachings about virtue, Protagoras settles on the interpretation of a poem in which he identifies an internal contradiction, challenging Socrates to defend the poet. Socrates interprets this poem to say that although virtue is vulnerable to misfortune, this misfortune must be the loss of knowledge, for virtue is knowledge; then, returning to the meta-conversation, he comments (perhaps on the grounds that such interpretations inevitably project our own beliefs onto the poem) that we do better to test our own beliefs.
6. Allowing for the possibility that some courage comes not from knowledge but from passion, natural and well trained?
7. On the grounds that Socrates is himself committed to ethical hedonism in the *Protagoras*, but criticizes it in e.g. the *Gorgias*, J. C. B. Gosling and C. C. W. Taylor, *The Greeks on Pleasure* (Oxford, 1982) argue that pleasure is long-term in the former, but sybaritic or present-aim in the latter;

- G. Rudebusch, *Socrates, Pleasure, and Value* (Oxford, 1999) argues that pleasure is modal (a pleasure of activity, in particular virtuous activity) in the former, but sensate in the latter. Because I believe the ethical hedonism of the *Protagoras* is a recently discovered position of the Many, I do not think this text gives us an answer at this level of philosophical detail. However, it's quite possible that Plato has a more fleshed-out account of the hedonism of the Many which can be reconstructed by considering the multiple contexts in which he discusses it.
8. The Many think this happens often even when one has knowledge but they do not say how often one has knowledge, contrary to the suggestion in Lombardo and Bell's translation (in Cooper and Hutchinson (eds), *Complete Works*): 'while knowledge is often present in a man, what rules him is not knowledge but rather anything else'. Compare the uses of *pollakis* at 352b and 353c.
  9. I translate the participles 'knowing' and 'being overcome' as not only reporting the condition of the agent (with '-ing') but as concerned with explaining the action ('although', 'because'). I've italicized these translation decisions to make them easier to spot.
  10. The context for *anti* is often exchange, for which Lombardo and Bell's translation, 'for the sake of', is unnecessarily intensional. In H. G. Liddell, R. Scott, and H. S. Jones, *Greek-English Lexicon*, 9th edn (Oxford, 1996) two passages are listed as having the sense 'for the sake of': Plato, *Menexenus* (237a) and Aristotle, *Nicomachean Ethics* (1110a21), but both are exchange contexts which needn't be read intensionally. In 'The Argument of Plato, *Protagoras* 351b–356c' (*Classical Quarterly*, 7 (1913), 100–4), J. L. Stocks argues that *anti* be understood as 'as compensation for', which suggests motive or purpose, and cites parallels in Plato *Phaedo* (69e) and *Lysis* (208e). It's true that if I take x at the price of y it's not a big leap to describe what I'm doing intensionally: I took x for the sake of y. But you can also describe my exchange non-intensionally; for example, if your interest is to evaluate the exchange, you might focus on the large magnitude of the negative value x and the small magnitude of the positive value y.
  11. Since Socrates *says* only that the views of the Many are absurd, not that they are internally inconsistent, one may argue for a different account of the absurdity, e.g. that Socrates is showing the Many to be absurd because they disagree with Socrates' obviously true psychological outlook, but I hope to show that Plato's achievement in the dialogue is far more ambitious and brilliant than simply to set out the differences between Socrates and the Many.
  12. G. Vlastos (ed.), *Plato: Protagoras* (New York, 1956), 'Introduction', xxxix. Vlastos does not agree with (what he takes to be) Socrates' judgement that it is absurd, inasmuch as it is not absurd for me to recognize that I ought to do one thing, and to want to do another (xlii–xliii).



13. R. Weiss, *The Socratic Paradox and Its Enemies* (Chicago, 2006) [= *Socratic Paradox*], says, 'how can good cause one to choose bad?', but then seems to hedge: 'What is absurd . . . is that someone would knowingly do what is bad because overcome by good . . . or knowingly do what is painful because overcome by pleasure when goods or pleasures are clearly . . . smaller/fewer than the bads or pains' (52–3).
14. D. Gallop, 'The Socratic Paradox in the *Protagoras*', *Phronesis*, 9 (1964), 117–29, at 121.
15. Irwin, *Plato's Ethics*, 84. Irwin calls the view Socrates attributes to the Many 'epistemological hedonism', eudaemonism along with the view that the good is pleasure and that judgements of pleasure are epistemologically prior to judgements about goodness (83). But in this case, Irwin must think Socrates attributes psychological eudaemonism, the thesis that all our actions are explained by our desire for the good, to the Many (cf. his discussion of psychological vs. rational eudaimonism, 52–5). Cf. Vlastos, 'Editor's introduction', in *Plato: Protagoras*, xxxix.
16. G. Santas, 'Plato's *Protagoras* and Explanations of Weakness', *Philosophical Review* (1966), 3–33.
17. And he needs this supposition because he needs a way of 'determining the relative strength of the conflicting desires independently of knowledge or information as to what action ensues from the conflict' (25). But why can't the action tell us which desire must have been stronger, with the strength of the desire then purporting to explain why the action took place? It's not a very informative explanation, but the explanation isn't being evaluated for its informativeness.
18. In 'Akrasia in the "Protagoras" and the "Republic"', *Phronesis*, 51 (2006), 195–229, M. Morris argues that hedonism is not essential to the argument, because (as he thinks) desires are evaluative beliefs. In his 'Introduction' to *Plato: Protagoras*, Frede says that it is easy to explain why Socrates thinks one can't act contrary to knowledge: that's because one can't act contrary to belief, for it takes partitioning of the soul to imagine how one can act contrary to belief (xxix–xxx); as a consequence, Frede also thinks hedonism inessential to the refutation of the Many (xxxi).

Among the few interpreters who do not suppose Socrates to be assuming the truth of psychological hedonism, M. Dyson, 'Knowledge and Hedonism in Plato's *Protagoras*', *Journal of Hellenic Studies*, 92 (1976), 32–45, points out that there is no mention of psychological hedonism in the refutation, that the statement of psychological eudaemonism at 358b comes too late for it to have been at work in the refutation, and that the gerundives at 356b–c cannot express psychological necessity; J. Clark, 'The Strength of Knowledge in Plato's *Protagoras*', *Ancient Philosophy*, 32 (2012) [= 'Strength of Knowledge'],

- 237–55, observes that Socrates only invokes psychological hedonism when addressing the sophists, after the conclusion of the refutation of the Many (244–5), and that his assuming it for the refutation would beg the question against the Many (246). But Clark vitiates these observations by arguing that Socrates attributes psychological eudaemonism to the Many (250). In fact, Socrates only invokes psychological eudaemonism when addressing the sophists, after the conclusion of the refutation of the Many (see Section 2.2.3).
19. M. Nussbaum makes sense of this dependence by showing that ethical hedonism plays the role of ensuring commensurability of value: *Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy* (Cambridge, 1986) [= *Fragility of Goodness*], 110.
  20. Irwin, *Plato's Ethics*, 53, explains the difference between psychological and rational eudaemonism as a difference between taking one's own happiness to be the ultimate explanation for one's actions as opposed to the ultimate justification for one's actions.
  21. The distinct issue of whether ordinary morality is eudaemonist (or egoist) is discussed in N. White, *Individual and Conflict in Greek Ethics* (Oxford, 2002), and L. Brown, 'Glaucón's Challenge, Rational Egoism and Ordinary Morality', in D. Cairns, F.-G. Herrmann, and T. Penner (eds), *Pursuing the Good: Ethics and Metaphysics in Plato's Republic* (Edinburgh, 2008), 42–60.
  22. In 'Gorgias and the Socratic Principle *Nemo Sponte Peccat*', *Journal of Hellenic Studies*, 77 (1957) [= 'Gorgias and the Socratic Principle'], 12–17, G. Calogero argues that Gorgias' *Defense of Palamedes* is the model for Socrates' defence speech in Plato's *Apology*.
  23. Of the two ways one might read the relationship between her reasoning and her anger here, namely (1) her anger opposes her reasoning, and wins out, or (2) her anger (rather than her love) commands her reasoning, I favour (2). Socrates must mean to reject both (1) and (2) in the case where knowledge is in the soul. That (2), and a Medea-like claim, are included in Socrates' target is given support by his description of the Many's position as being that often, when knowledge is present in someone, it does not rule but what rules is rather *at one time* anger, *at another*, pleasure, or pain, or love (352b).
  24. Cf. *Homeric Hymn to Aphrodite*, and Helen's return to Paris under the influence of Aphrodite in *Iliad* 3. One might think it makes a difference whether the phenomenon at issue is conceived of as one thing driving out another as opposed to one thing winning over another while both are present in the soul, but my examples from Euripides don't discriminate between these. What I'm illustrating from Euripides is the conception of action as the outcome of whatever element in the soul has the most strength, where strength is not being conceived of in evaluative terms—whether the stronger acts by driving out or by conquering the weaker. Interestingly, C. Bobonich

finds the latter picture compatible with the *Protagoras* in his 'Plato on *Akrasia* and Knowing Your Own Mind', in C. Bobonich and P. Destree (eds), *Akrasia in Greek Philosophy from Socrates to Plotinus* (Leiden, 2007), 41–60.

For the record, Gorgias is no principled psychological eudaemonist, but in his *Encomium of Helen* argues that Helen is blameless for running off with Paris, whether the cause of her doing so was Fate, the will of the gods, necessity, force, persuasion, or love (6); when it comes to love's form of compulsion, Gorgias notes that fear, which is one of the things that come into us through sight (love is another), drives out thought (*apesese kai exêlasen ho phobos to noêma*, 17).

25. I'm grateful to Tad Brennan for helping me to clarify the presentation of this argument.
26. As is claimed by C. C. W. Taylor (trans. and comm.), *Plato: Protagoras* (Oxford, 1976) [= *Protagoras*], 180–1.
27. D. Wolfsdorf, in 'The Ridiculousness of Being Overcome by Pleasure: *Protagoras* 352b1–358d4', *Oxford Studies in Ancient Philosophy*, 31 (2006) [= 'Ridiculousness'], 113–36, argues that Socrates identifies the following contradiction in the Many's account of *akrasia*: being overcome by goodness (which is the same as pleasantness) implies that the goodness of the action taken is greater than that of the alternative; however, by definition, the goodness of the action taken is less than that of the alternative. In conceiving of goodness and pleasantness only in terms of quanta, Wolfsdorf's interpretation seems pretty close to mine, but I take the Many's claim that what is overcome is knowledge to be crucial, since I rely on knowledge's being factive to commit the Many to the result that the goodness of the action taken is less than the alternative action not taken, whereas Wolfsdorf (who thinks there is no important difference between knowledge and belief in this argument), relies only on the Many's view that the action taken is mistaken.

In contrast to my interpretation, which locates the Many's absurdity in their conflicting claims about relative magnitudes, A. Callard, in 'Ignorance and *Akrasia* in the *Protagoras*', *Oxford Studies in Ancient Philosophy*, 47 (2014), 31–80, argues that Socrates is showing the apparent akratic to lack self-knowledge: the akratic thinks that the pleasure by which she is moved is an appearance, but in fact she believes that it is better, the evidence that this is her belief being her action; and in fact her so-called knowledge that the other pleasure is better is actually just an appearance. This interpretation not only attributes several undefended psychological views to Socrates and has him foisting these views on the Many (e.g. that only beliefs and knowledge, not appearances, can motivate actions), but also restricts Socrates' refutation to the Many's explanation of their *own* akratic actions.

As evidence for her assumption that appearances can't motivate, Callard cites *Republic* 505d–e, where Socrates says that everyone wants not what seems, but what really is, good. But that text says many people are content to have what *seems* fine or just (*ta dokounta*), so it doesn't seem to be a general truth about motivation that an appearance can't motivate, and in any case, what Callard needs is not a contrast between our desire for what is rather than what seems, but a contrast between belief and appearance. Callard claims that in the *Protagoras* an appearance does not motivate unless one believes it is veridical, but Socrates' point about how the measuring art disables visual appearances of size from influencing our actions in the *Protagoras* is weaker: it is that an appearance one knows *not* to be veridical doesn't motivate.

28. Taylor wonders whether the objector is proposing to modify Socrates' account by adding propinquity as a factor in determining the value of pleasures or objecting that our psychological preference for the present explains our choice of the lesser pleasure. He finds Socrates' reply 'disappointing' in either case, amounting to either mere rejection of the modification or mere assertion that the psychological preference for the present must itself be the result of mistaken assessment (*Protagoras*, 188–9). (NB Socrates switches from '*ta*' to '*to*' at 356b2.) Wolfsdorf says that Socrates 'addresses' this objection, but it seems that on Wolfsdorf's account ('Ridiculousness', 123–6), Socrates 'acknowledges' that the proximate pleasure can seem smaller than the distant, but denies that it is. But if the Many are offering proximity as an explanation of how something apparently larger can overpower something actually larger, what is Socrates doing to address it?
29. 'Pointless', says Taylor, and goes on to argue that the passage gives an account of correct choice to prepare the Many for Socrates' 'error' account of incorrect choice (*Plato: Protagoras*, 190). But his translations, 'have to take' and 'have to do', and Lombardo and Bell's translations 'must' and 'have to', blur the issue by suggesting psychological necessity.
30. Cf. Wolfsdorf: in context these endings indicate prudential obligation. However, Wolfsdorf casts unnecessary doubt on the generality of the point ('Ridiculousness', 124–5) by citing *Gorgias* 499e ('aren't the good pleasures and pains both to be chosen (*haireteon*) and to be done (*prakteon*)?') and 506e ('the pleasant is to be done (*prakteon*) for the sake of the good') as cases of psychological necessity. Wolfsdorf thinks this is based on the 'psychological axiom' that 'all things must be done (*prakteon*) for the sake of good things' (499e), which refers back to *Gorgias* 468b ('it's for the sake of the good that all those who kill, exile, etc. do so'); Chapter 3 corrects this (widespread) misunderstanding of that passage. Clark ('Strength of Knowledge', 243) notes that the *-teos* endings are embedded in clauses following the imperative

'say', which suggests 'ought'. Of course, H. W. Smyth's *Greek Grammar* (Cambridge, MA, 1956) (s. 358) says that *-teos* endings express necessity, and suggests the translation 'must', but this is the 'must' of hypothetical necessity, so Socrates' point is: 'given the end of living well, one must (if one wants to achieve this end) weigh pleasures against pleasures and take the greater'.

31. Perhaps it also speaks to Protagoras' relativism, which otherwise makes only a brief showing (334a–c): rather than being the measure of truth, appearances only have power in the presence of ignorance and are rendered powerless by the art of measurement.
32. I take this list from R. Singpurwalla, 'Reasoning with the Irrational: Moral Psychology in the *Protagoras*', *Ancient Philosophy*, 26 (2006) [= 'Reasoning'], 243–58 at 247.
33. Strictly speaking, the Many had thought that other motivations apart from pleasure and pain could also overrule knowledge: desire, love, fear (352b–c), and one might add, anger. Among these, pleasure is special, because it is both a feeling or force and a value. Socrates' refutation wouldn't work so beautifully against, for example, 'overcome by anger'. However, even though the refutation wasn't of the general claim that any passion can overcome knowledge, the alleged phenomenon of knowingly choosing the worse because of pleasure is thought to invalidate Socrates' idea that knowledge is always the master in the soul, so that showing the alleged phenomenon doesn't occur gives particular interest to Socrates' idea, and opens the door for a psychology that would explain how knowledge is the master. Socrates could also reduce other passions to pleasure and pain, and in any case, the new psychology he is about to develop will define the other passions in terms of beliefs about pleasure/good and pain/bad.
34. (E4) is much more circumspect than (E1), allowing that if the agent's belief that an action A is the best possible doesn't result in her doing A, she may be acting unwillingly. According to T. Penner, 'no-one acts contrary to his belief about what's best' (= the consequent in (E1)) is a 'direct corollary' of 'no-one does wrong willingly' (= E4). Penner asks the excellent question why, given (E1) and (E4), which he supposes that Socrates assumes all along, Socrates *argues* against the Many that no one acts contrary to their knowledge of what is best—for if you think of knowing *p* as just a case of believing *p*, there is no need of argument. Penner's explanation of why Socrates argues that knowledge can't be overcome is that Socrates wants to show that knowledge guarantees not only that we do not act contrary to our belief about what is best in the moment (there is no 'synchronic belief *akrasia*'), but also that we do not change our beliefs under the influence of appearances (there is no 'diachronic belief *akrasia*'), in 'Plato and Davidson on the Parts of the Soul',

- 46–8; T. Penner, 'Knowledge vs. True Belief in the Socratic Psychology of Action', *Apeiron*, 29 (1996), 121–69; T. Penner, 'Socrates on the Strength of Knowledge: *Protagoras* 351B–357E', *Archiv für Geschichte der Philosophie*, 79 (1997) [= 'Socrates on the Strength of Knowledge'], 117–49. Section 2.1 has provided ample reasons against supposing that Socrates can simply assume that there is no synchronic belief *akrasia*.
35. N. Gulley reasons: if a person does wrong, he must be ignorant (lacking knowledge); to be ignorant is to have a false belief; so a person can't act contrary to a true belief, in 'Socrates' Thesis at *Protagoras* 358B–C', *Phoenix*, 25 (1971), 118–23 at 120. J. Allen agrees: the statement identifies ignorance and false belief, on the one hand, and knowledge and true belief, on the other: see 'Dialectic and Virtue in Plato's *Protagoras*', in B. Reis (ed.), *The Virtuous Life in Greek Ethics* (Cambridge, 2006) [= 'Dialectic and Virtue'], 6–31 at 24. Cf. *Republic* 382b for the association of ignorance and false belief about the most important things.
  36. G. Vlastos, 'Introduction', in Vlastos (ed.), *Plato: Protagoras* (New York, 1956), xxxix. But elsewhere Vlastos argues that the point about belief is just that no one can act contrary to beliefs based on knowledge: G. Vlastos, 'Socrates on *Akrasia*', *Phoenix*, 23 (1969), 71–88.
  37. Perhaps 'in human nature' qualifies the claim: it may belong to non-human animal nature to do something other than what seems best.
  38. In claiming that the early dialogues show Socrates recognizing non-rational desires, Brickhouse and Smith follow Devereux's 'Socrates' Kantian Conception of Virtue', 395. I do not find their evidence for non-rational desires in the early dialogues compelling. Among the passages Devereux cites are *Laches* 191d–e, which says that courageous people are *deinos* in battle against fears, desires, and pleasures; *Gorgias* 507b, which says the temperate man must also be brave for he resists and endures pleasures and pains where he should; *Gorgias* 493a–b, which says that there is an element or a region in the soul in which the [pleasure-seeking] appetites lie. Brickhouse and Smith add several more passages, such as *Apology* 21c–22a, 24a, where Socrates says people hate him for showing up their ignorance; and *Apology* 29e–30e, where he says he tries to shame people. The problem with such evidence is that fear, desire, endurance, appetite, hatred, and shame can all be given *Protagoras*-style accounts in terms of what appears good or bad to the agent (see Section 2.3).
  39. Singpurwalla, 'Reasoning', takes irrational desires to be, like the passions, false beliefs about good and bad. Her account makes sense of how knowledge might make appearances lose their power and 'give us peace of mind' (356d): if an irrational desire/passion is a judgement about good or bad, then knowing that disgrace is worse than death will keep me from fearing death more than disgrace. Death might continue to appear quite terrible, but that appearance

cannot lead me to form the judgement that death is more terrible than disgrace—for that judgement would conflict with my knowing, and on that basis judging, that disgrace is more terrible than death. However, Singpurwalla also posits a cause for the irrational desire/false belief: 'attraction', which explains the phenomenon of pleasures appearing larger because they are closer. But false beliefs based on visual appearances will then have a completely disanalogous explanation.

40. Brickhouse and Smith, *Socratic Moral Psychology*, 73 n. 5, their emphasis.
41. Brickhouse and Smith, *Socratic Moral Psychology*, 79.
42. Allen, 'Dialectic and Virtue', 7.
43. I owe this point to Tad Brennan.
44. Irwin, *Plato's Ethics*, 52–5, seems to conclude that in the *Euthydemus* Socrates affirms only rational eudaemonism, failing to distinguish it clearly from psychological eudaemonism.

# 3

## Why is Wrongdoing Unwilling?

### 3.1 Introduction

In Chapter 2, I argued that the psychological eudaemonist thesis that we always do what we believe is the best of the things we can do is a hypothesis introduced in the *Protagoras* by Socrates to explain how Protagoras might guarantee that the *technê* he teaches makes his students good, for it implies that when we do things that are worse than other things we could do, it must be due to ignorance of their relative badness. The hypothesis also secures Socrates' thesis that virtue is knowledge. However, because the plausibility of this hypothesis depends on the refutation of the Many's account of knowledge being overcome by pleasure, which in turn depends on the identification of the good with pleasure, which Plato has Socrates reject everywhere else, we should not be too quick to assume its presence in other dialogues. In the present chapter I argue that the 'Socratic intellectualist' thesis 'no one does wrong willingly' is based not on psychological eudaemonism, but on the more fundamental thesis that human nature seeks, and so we engage in purposive action in order to secure, our real good. Attention to the dialectical contexts in which 'no one does wrong willingly' appears in the dialogues reveals that Plato's arguments for this thesis in fact adopt and develop two sophistic ideas: (i) there is such a thing as real/natural advantage, which it is in our nature to desire and pursue, and (ii) that which we do or undergo contrary to our natural desires and pursuits we do or undergo unwillingly (*akôn*).

The chapter is organized as follows: Section 3.2 argues that psychological eudaemonism is not a sufficient basis for 'no-one does wrong willingly'. Section 3.3 discusses sophistic claims about the conditions that make actions unwilling: compulsion, and the sort of ignorance that



would prevent the agent from formulating the relevant intention and so calls into question the attributability of the action as described to the agent. Section 3.4 shows how the *Gorgias* draws on the characterization of compulsion in order to distinguish willingness and unwillingness *in the case of actions undertaken as means for the sake of some (further) end*. On the assumption that unjust actions are undertaken for the sake of our happiness, which we naturally desire, Socrates maintains that wrongdoing is unwilling because it makes us unhappy. In the absence of this means–ends framework, the *Hippias Minor* gives the apparently conflicting result that the good man does wrong willingly. Section 3.5 shows that across the dialogues, including those that recognize that we sometimes act contrary to what we judge to be best, contrariety to our conative pro-attitude towards our actual good is what makes something (action, experience, character-state) unwilling. So Section 3.2 demonstrates that psychological eudaemonism isn't sufficient for 'no one does wrong willingly' and Section 3.5 that it is not necessary. Finally, the Appendix shows how this chapter's account of 'no one does wrong willingly' in Plato illuminates Aristotle's arguments in *Nicomachean Ethics* III.5 for why we do both good and bad actions willingly and are both good and bad willingly.

## 3.2 Why Psychological Eudaemonism Is Not the Basis for 'No One Does Wrong Willingly'

To motivate my alternative to psychological eudaemonism as the basis for 'no one does wrong willingly', this section draws attention to the weaknesses of arguing to it from the assumption of psychological eudaemonism. Here is a reconstruction of how the reasoning from psychological eudaemonism to 'no one does wrong willingly' is supposed to go:

- (R1) Everyone does/wants to do what he believes to be good/best (= (most) beneficial to him).
- (R2) Therefore, pursuit of (really) bad (= harmful) things is due to ignorance of their badness.
- (R3) Actions done due to ignorance are unwilling.
- (R4) Therefore, no one does bad things willingly.
- (R5) Agents do wrong for the sake of what they believe to be good/beneficial.

- (R6) Wrongdoing is (not good/beneficial but) bad/harmful.
- (R7) Therefore, wrongdoing is due to ignorance (either that the action done is wrong or that it is bad).
- (R8) Therefore, wrongdoing is unwilling.<sup>1</sup>

(R1)–(R4) are the core of the argument; (R5)–(R8) just apply the conclusion about the involuntariness of bad actions (R4) to wrongdoing on the grounds that wrongdoing is bad. To see the problem with (R1)–(R4), consider: if I ignorantly pursue something that is in fact bad for me (R2) because I believe it is good for me (R1), I am acting on my desire, which is for what I believe is good, or to do what I believe is good (R1), so what is unwilling about my action (R3)? For example, suppose I desire and pursue an MBA because I believe it is good for me (R1). In fact, an MBA only narrows my horizons, so I must be ignorant of this badness for me (R2). How does my ignorance of the badness of the MBA make my action in pursuit of it unwilling (R3)? It cannot be because it contravenes my desire (R1), since my desire was to do what I believed good, namely to get the MBA and I am doing that. Nor can it be because of my ignorance of the real effects of an MBA alone. For suppose I study philosophy in order to get into a good law school, because I believe law school is good for me (R1), ignorant that the study of philosophy in fact ennobles students' interests. Is my benefiting from the study of philosophy in this way unwilling because it is due to ignorance (R3), or because it is not in conformity with my desire (R1)? But this would be an odd thing to say. One could argue that the standard for willingness is not conformity to my actual desire, but to what my desire would be if I were fully informed, but it would take some argument to establish this as the standard for acting willingly. And if this is the standard of willingness, then it isn't the psychological eudaemonism of (R1) and (R2) but what I would want if fully informed that grounds (R3).

A further problem with grounding 'no one does wrong willingly' in psychological eudaemonism is that Plato does not seem to accept (R3), at least not by itself. In the *Apology*, Socrates treats one kind of ignorance as a condition for acting willingly, and another kind as a condition for acting unwillingly. He argues:

- (A1) The vicious harm their associates while the good benefit theirs (25c).
- (A2) If I corrupt my associates, then I make them vicious (25d).

- (A3) No one wants (*bouletai*) to be harmed (25d).
- (A4) If I corrupt my associates willingly (*hekonta*), I must (by A3) be ignorant of (A1) and (A2) (25e).
- (A5) It is not plausible that I am ignorant of (A1) and (A2) (25e).
- (A6) Therefore, either I do not corrupt my associates, or if I corrupt them, I do so unwillingly (26a).
- (A7) If I corrupt my associates unwillingly, I should be instructed and exhorted, not taken to court (26a).

According to (A4), ignorance of the fact that corrupting one's associates makes them vicious, and so such as to harm one, is what would make it possible for one to corrupt them *willingly*. Why would ignorance make it possible to perform an action willingly? If we understand *hêkon* in (A4) as 'intentionally' and *bouletai* in (A3) as 'intends to' (which are standard senses of the Greek terms),<sup>2</sup> the reasoning would go as follows: if Socrates corrupts his associates intentionally, he must intend to corrupt them, but if he intends to corrupt them, he must intend the (obvious) consequence (A5), that they become harmful to their associates, of which he is one, which is tantamount to intending that he be harmed (A1, A2). But no one intends to be harmed (A3).<sup>3</sup> In this argument, even though the ignorance in question (that harm to oneself is the obvious consequence of corrupting one's associates) turns out not to be possible (unless one is mad), it would, if possible, have been a condition under which Socrates' action could be willing. So in the *Apology* the impossibility of Socrates'/a sane person's doing bad willingly (intentionally) is secured by the implausibility of a certain ignorance (A5), not by pronouncing every action done due to ignorance 'unwilling', as on (R3). Further, the other alternative (A7), that Socrates corrupts his associates unwillingly (unintentionally, perhaps intending instead to inquire into virtue jointly with them), must also involve ignorance, for what other than ignorance would be remedied by instruction and exhortation? It would be quite reasonable for Socrates to think that ignorance of something so counterintuitive—that inquiring into virtue jointly with other people (as Socrates does) corrupts them—is the sort of ignorance that calls for instruction rather than punishment.<sup>4</sup> So the argument of the *Apology* calls for distinguishing between the unwilling-making and not-unwilling-making kinds of ignorance, rather than for (R3). And it makes no use of (R1) at all. (A3), 'No one wants to be harmed', doesn't take it that the harm that no one wants or intends is what

the agent believes is harmful, but rather what is in fact harmful—which is what makes the agent's ignorance relevant in the first place.<sup>5</sup> So while (R3) might explain why someone acted self-destructively as they did, (A3) is required to explain why in doing so they acted unwillingly.

### 3.3 Who Believes 'No One Does Wrong Willingly'?

In the *Protagoras*, Socrates interprets a poem by Simonides which says:

- (S1) Simonides won't blame a man so long as the man is not bad or too hardened, and is supportive of his city's justice.
- (S2) Simonides praises and loves those who do nothing shameful willingly.
- (S3) Doing only good things even through great misfortunes (which the gods sometimes inflict on mortals) is a perfection possible only for the gods (339a–347a).

As Socrates interprets the poem, 'willingly' in Simonides' (S2) must attach to 'praise and love': 'willingly do I praise and love him who does nothing shameful', for Simonides, who is one of the wise, can't mean that he praises and loves those who do nothing shameful willingly, since (according to Socrates) none of the wise believe that anyone does wrong willingly (345d–e: *oudeis tôn sophôn andrôn hêgeitai oudena anthrôpôn hekonta examartanein oude aischra te kai kaka hekonta ergazesthai, all' eu isasin hoti pantes hoi ta aischra kai ta kaka poiountes akontes poiousin*<sup>6</sup>). This is manifestly not what Simonides means, since his whole point is that in view of the impossibility for human beings of doing only good things (S3), he judges human beings by asking whether they do bad things, which they inevitably do as a result of misfortune (*sumphora*), willingly or only unwillingly.<sup>7</sup> (Confirmation that (S2) cannot mean what Socrates says is that attaching 'willingly' to 'praise and love' instead of to 'do wrong' compromises the metre.<sup>8</sup>) Indeed, Socrates' interpretation seems to follow the principle, 'attribute to Simonides what seems true and then somehow get that out of his words'. But Simonides himself is clearly saying that *some* wrongdoing is unwilling, due to misfortune, and so not blameworthy, by contrast with willing wrongdoing, which is cause for blame. In another poem which also seems to conclude that it is impossible for a human being

always to be good (*PMG* 541), Simonides enumerates, among the things that force (*biatai*) a person to act unwillingly (*ækonta*): irresistible desire for gain, Aphrodite's overpowering sting, and the love of victory. This raises questions: if greed, lust, and ambition can render the action they motivate unwilling, can actions whose badness seems to consist in their being greedy, lustful, or ambitious ever be willing? Or, if we think the badness of actions consists in the harms they cause, aren't greed, lust, and ambition just the sorts of motives that lead to bad actions? How is a bad action ever willing? Socrates' interpretation of Simonides seems to be on to something, even if it isn't what Simonides had in mind.

Nevertheless, Socrates' claim that none of the wise believe (*hêgêtai*) that anyone does wrong willingly seems *prima facie* false, since plenty of the wise allow that people sometimes err, and do bad things, and shameful things, willingly. For example, in the *Hippias Minor* (discussed further in Sections 3.3.1 and 3.4.3), Hippias judges Achilles a better man than Odysseus on the grounds that Odysseus spoke falsehoods willingly but Achilles unwillingly. In Aeschylus' *Prometheus Bound*, Prometheus says of his theft of fire to save the human race in defiance of Zeus, "willingly, willingly I did wrong" (266). In Euripides' *Iphigeneia in Aulis*, Menelaus says Agamemnon willingly sent for Iphigeneia to be sacrificed (360; although one could argue that doing so was not wrong, as Calchas had prophesied that he must do this for the Achaeans to be able to sail to Troy, and Agamemnon was under oath to help Menelaus retrieve Helen), and Achilles says Odysseus was chosen by the army, but willingly (*hairetheis hekôn*), to capture Iphigeneia for slaughter (1364; however, Odysseus too was under the same oath). Even if Hippias, Achilles, and Menelaus shouldn't be counted among the wise, Prometheus surely should be. So perhaps Socrates means that the wise are *committed* to 'no one is bad or does wrong willingly'. They may not avow it, or ever occurrently believe it, but perhaps it is an implication of their avowals or occurrent beliefs. To see what could commit the wise to 'no one does wrong willingly', we should begin by surveying the contexts in which they claim that wrongful actions are unwilling.

### 3.3.1 Unwillingness in Excusing Arguments: Compulsion

In contexts of aretaic assessment (e.g. in answer to the question, 'who is the best of the Achaeans?'), *excusing arguments* maintain that a given

wrongful act was compelled by some force not in a human being's power to resist, and consequently that the agent being assessed should not be praised or blamed for having done it.

The *Hippias Minor* offers a clear case of appeal to unwillingness due to compulsion to make an excusing argument. Hippias defends his opinion that Achilles is the best of the Achaeans and better than wily (*polutropos*) Odysseus, even though Achilles too speaks falsely, on the grounds that Achilles speaks falsely not from a plan but unwillingly (*ouk ex epiboulês . . . alla akôn*, 370d–e). Although Achilles had announced his intention to leave Troy, it was because he was compelled (*anangkastheis*) by the plight of the army to help that he did not carry out his intention (370e); when he said different things to Ajax than to Odysseus, it was out of simple-mindedness or good intentions;<sup>9</sup> by contrast Odysseus planned (*epibouleusas*) to deceive (371d–e). In support of his opinion, Hippias appeals to the law, which is more lenient on the unwilling wrongdoer, who does injustice unknowingly (*mê eidôs adikêsêi*), than on the willing (372a). So on Hippias' view, a willing falsehood is relevant to assessing the character of the one who tells it; not so an unwilling one. The account of Achilles' unwilling falsehood is that some sort of (presumably non-culpable) weakness—for example, ignorance, folly, or infirmity—in Achilles renders him unable to master the circumstances—that is, the plight of the army. By contrast, liars are powerful (*dunatous*, 365d7) rather than weak; they speak falsehood from intelligence and cunning (*phronêseôs kai panourgias*, 365e4–5) rather than ignorance or folly; they know what they are doing (*epistantai hoti poiousin*, 365e7), and are wise (*sophoi*, e10); they are able to speak falsehood when they want to (366b).<sup>10</sup>

Outside the text of Plato, Gorgias' showy and challenging *Encomium of Helen* takes such excusing arguments to their limit. The *Encomium* praises Helen and argues that she is blameless for running off with Paris whether she did so because of the will of fate or of the gods or the decrees of necessity (*anangkê*), because she was taken by force, because she was persuaded by speeches, or because she was captured by love, for each of these is an irresistible force that comes into the agent from the outside, against which she would have been powerless (1–6).<sup>11</sup> Gorgias begins with fate, the gods' plans, and necessity, which everyone agrees a human cannot resist; he then likens force to necessity in being irresistible; speech too turns out to have irresistible power, just like physical force despite its invisible and tiny body, as well as 'divine' effects (8–15); finally, sight

is passively affected by the nature of whatever it sees, and love is sight's natural and necessary response to the beautiful, so it follows that love either is a god and so cannot be resisted by a human being, or is a sickness that should be regarded as a misfortune (15–19). Gorgias' argumentative strategy, of assimilating cases in which Helen's attitudes were the cause of her doing as she did to clear cases of compulsion, in which an external agency caused Helen to do as she did, makes inescapable the question suggested to us by Simonides, namely whether all wrongdoing is unwilling. Indeed Gorgias' *Encomium* raises the question how any action could be willing. If even persuasion and desire compel or necessitate and so excuse the actions they result in, what uncompelled cause of action could there be? Can an agent ever be responsible for her actions?

A closer look at what Gorgias says about the power of speech (*logos*) and persuasion indicates the contrast with willing or uncompelled action on which Gorgias is relying:

If speech (*logos*) persuaded her and deluded her mind, even against this it is not hard to defend her or free her from blame, as follows: speech is a powerful master and achieves the most divine feats with the smallest and least evident body . . . To its listeners poetry brings a fearful shuddering, a tearful pity, and a grieving desire, while through its words the soul feels its own feelings for good and bad fortune in the affairs and lives of others . . . Sacred incantations with words inject pleasure and reject pain, for in associating with the opinion of the mind, the power of an incantation enchants, persuades, and alters it through bewitchment. The twin arts of witchcraft and magic have been discovered, and these are illusions of mind and delusions of judgement. If all men on all subjects had memory of the past, <understanding> of the present and foresight into the future, speech would not be the same in the same way; but as it is, to remember the past, to examine the present, and to prophesy the future is not easy; and so most men on most matters make opinion an advisor to their minds . . . What reason is there, then, why Helen did not go just as unwillingly<sup>12</sup> under the influence of speech as if she were seized by the violence of violaters? For persuasion expelled her thought (*nous*)—persuasion, which has the same power, but not the same form, as compulsion (*anangkê*) . . . To see that persuasion, when added to speech, indeed molds the mind as it wishes, one must first study the arguments of astronomers, who replace opinion with opinion: displacing one but implanting another, making incredible, invisible matters apparent to the eyes of opinion. Second, compulsory debates with words, where a single speech to a large crowd pleases and persuades because it is written with skill (*technê*), not spoken with truth. Third, contests of philosophical arguments, where the speed of thought

makes it easy to change a conviction based on opinion. The power of speech has the same effect on the disposition of the soul as the disposition of drugs on the nature of bodies . . . [I]f she was persuaded by speech, her fortune was evil, not her action . . . (8–15)<sup>13</sup>

Persuasive speech has power over its audience due to the audience's infirmity of memory, understanding, and foresight (11); given these infirmities, the audience relies on opinion, which it is the power of persuasive speech to produce and replace (13). Although Gorgias cites the power of poets, priests, astronomers, and philosophers to drive out and implant opinions in their audiences at will, it is not truth, but their skill at producing feelings, at pleasing, at making invisible things apparent, and at arguing quickly, that enables them to do this. Gorgias' account of persuasion, along with his contrast between the compulsive character of love, on the one hand, and the plans of thought (*bouleumasin gnômês*) and provisions of skill (*paraskeuais technês*) (19) on the other, suggests that the power of persuasive speech is correlated with the powerlessness of the soul it influences, and that resistance to persuasive speech requires some power in the audience (e.g. thought, skill, memory, understanding, and foresight). So Gorgias leaves open the possibility that Helen's wrongdoing would be uncompelled or willing if done from some power in Helen or of Helen; however, his speech can safely ignore these possibilities, for who would suppose Helen to be equipped with any power rather than weakness?

Although persuasive speech is just one of the possible compellers Gorgias considers, alongside the gods, necessity, and love, it occupies the central third of the *Encomium*. Gorgias' audience can hardly fail to register that his speech is doing to them what it says Paris' speech would have done to Helen and other instances of speech do to their audiences. Thus, the *Encomium* is itself a paradigm of the power of persuasive speech: if we believe Helen acted out of love or because she was persuaded, we believe she is to blame for her actions, but Gorgias' speech challenges us to articulate the difference between these motivations and the compulsion exercised by abduction or the will of the gods or fate—failing which we are compelled to agree with Gorgias that Helen is not to blame. And as we the audience are compelled by the power of persuasive speech, Gorgias the possessor of the art of persuasive speaking exercises power over us.



### 3.3.2 *Unwillingness in Immoralist Arguments: Contrariety to Nature*

In the *Republic*, Glaucon reports the argument that all who practise justice do so unwillingly (*akontes*), as something necessary rather than good (358c1–4), and from a lack of power to do injustice with impunity (359c1–2). Every nature desires to outdo others and get more and more, but is led by law and force to honour fairness (359c3–6). In other words, our just behaviour is unwilling because justice and law compel us to act contrary to our natural motivations. In the *Gorgias*, Callicles faults conventional justice (or ‘the justice of law’) for frightening and shaming those capable of a greater share into taking only an equal share; he favours the ‘natural justice’ that allows the superior to have the greater share of which they are capable (483a–e). These passages conceive of our nature as leading us to desire and pursue our good. The *Republic* passage describes justice, something that frustrates our natural desire, as making our actions ‘unwilling’; the *Gorgias* passage conceives as ‘natural’ the exercise of capability in the pursuit of the greater share (which we naturally desire), and characterizes conventional justice as compelling, indeed, ‘enslaving’ the naturally capable.

In *Truth*, Antiphon vividly illustrates this idea of our nature as involving internal impulses directed towards our advantage, which the law constrains

[R]ules have been made for the eyes, what they should and should not see, and for the ears, what they should and should not hear, and for the tongue . . . and for the mind, what it should and should not desire . . . [but] the things from which the laws dissuade us are in no way less congenial or akin to nature than the things towards which they urge us . . . The advantages laid down by the laws are bonds on nature, but those laid down by nature are free.<sup>14</sup>

To Gorgias’ conception of compulsion as due to a power in the compelling party that acts irresistibly on a weaker party, these passages add the thought that the compelling power frustrates or impedes some natural impulse in the compelled.

It may seem that such statements that no one willingly does just things contradict Socrates’ claim in the *Protagoras* that none of the wise believe that anyone willingly errs or does bad or shameful things, but the contradiction is superficial, and a distinction between what the wise say and what they are committed to allows us to see how Plato resolves the

contradiction. As Gorgias' *Encomium* shows, the excusing arguments based on compulsion actually made by the wise seem to have no principled reason to restrict the judgement 'compelled/unwilling' to some wrongful actions rather than extending it to all actions. The arguments from nature suggest a way to restrict the judgement 'unwilling' to actions that are harmful to the agent, for if the action is advantageous, it is in line with the agent's natural impulses, and so not compelled or unwilling. What remains at issue is whether wrongful or shameful actions are bad for the agent, and just and fine actions good, or vice versa.<sup>15</sup>

### 3.3.3 *Ignorance in Tragedy and the Law*

In the contexts considered so far, we have seen that compulsion makes actions unwilling, and in Gorgias' *Encomium*, we have seen that ignorance is a condition that makes compulsion possible. I argue below that in tragedy, too, ignorance renders actions unwilling not by itself but only because it enables one person's agency to be hijacked by another. In the law, while it seems that some kinds of ignorance do, by themselves, make the action unintentional, a sophistic refinement suggests that these kinds of ignorance call into question what is the correct description of an unintentional action.

For example, in Euripides' *Hippolytus*, Artemis consoles Theseus, who has put a death-curse on his son Hippolytus because he believed his wife Phaedra's false accusation that Hippolytus raped her, 'for you killed [your son] unwillingly' (*akôn gar ôlesas*, 1433). Theseus' action is unwilling not simply because he acted out of ignorance, but also because he was Phaedra's and ultimately Aphrodite's pawn; Artemis says that when the gods so ordain, humans are to be expected to make disastrous mistakes. Similarly, when Heracles kills his wife and children having only just saved them, because Hera sends Madness upon him and he does not know what he is doing (*Heracles* 900–80), and when Agavê decapitates her son Pentheus while in a Dionysian frenzy, mistaking him for a lion cub (*Bacchae* 1173–1215, 133–42, 1251–79), the hand of the gods in what is done overdetermines the judgement 'unwilling'.<sup>16</sup> And in Sophocles' *Women of Trachis*, Deianira, who poisons Heracles, thinking she is giving him a love-potion the centaur Nessus once gave her, is said to have killed Heracles, and erred (*hêmarten*) unwillingly (727–8, 935, 1123)—but once again, the judgement that her action was unwilling seems to be due not simply to her ignorance, but also to the fact that Nessus, who

sought revenge on Heracles, has used her. Finally, in Sophocles' *Oedipus at Colonus*, when Oedipus says Creon should not blame him, for he killed his father and wed his mother not knowing who he was that was doing it, or to whom (*mêden sunieis hôn edrôn eis hous t' edrôn*, 973–6), it may seem that at least here, ignorance suffices for unwillingness, for there is no other agent to identify as the source of the action. But again, the fact that—as Oedipus complains—all that happened was fated even before he was conceived (960–99) overdetermines the unwillingness of his actions.

One might expect ignorance to function as an unwillingness-maker in legal contexts, especially since the law treats intentional and unintentional actions differently. But such actions raise the question of how to correctly describe the action for which the agent is said to be responsible, suggesting that the action under the description which is said to be unwilling is not even attributable to the agent. For example, Antiphon's Second Tetralogy presents speeches from the prosecution and defence in a (fictitious) legal case of a youth who, while practising javelin-throwing, hit a boy who was running out to collect fallen javelins, and was charged with involuntary homicide (the penalty for which would be a year's exile). Antiphon's defence argues that according to the precise account (*akribos logos*) of what happened, the victim was responsible for the death, since his running out into the path of the javelin, and not the throwing of the javelin, was the event responsible for his death—there was lots of other javelin-throwing going on, and none of those throws were responsible for the victim's death.<sup>17</sup> The issue of what the javelin-thrower's action was arises here because while in the normal case action-description may be determined by the agent's (likely) intention (as at *Apology* 25c–26a), this is not possible in the case of unintentional action. Antiphon seems to be suggesting that for the case of unintentional action, usual cause-and-effect relations should determine the description under which an action is attributed to an agent: javelin-throws don't usually result in hitting-people/killing-people, but runnings-in-the-path-of-javelins do. So what the youth did was not the action, '[unintentionally] hitting/killing the boy with his javelin' but rather, '[intentionally] throwing the javelin', whereas what the boy did was, '[unintentionally] running in the path of the javelin'.<sup>18</sup>

So far, we have found that actions are counted unwilling for two distinct kinds of reasons: (1) because due to a power acting on the agent

that is both contrary to the agent's natural (self-interested or good-directed) impulses and difficult to resist; in this case ignorance can function as one of the conditions of weakness that enable outside forces to exert their power on an agent. And (2), because due to ignorance of the circumstances of action so that the agent cannot be said to have intended the action—which problematizes the description under which an action may be attributed to the agent.<sup>19</sup>

### 3.4 Instrumental Desires and the Unwillingness of Wrongdoing

Plato's *Gorgias* begins with Gorgias making claims that seem to recapitulate his *Encomium* case for the compelling power of persuasive speech: Gorgias boasts that his expertise in persuasion, rhetoric, is the source of freedom for oneself and rule over others; it enables its possessor to have such experts as the doctor and the physical trainer as his slave (452d–e). To be free oneself and to rule over others are paradigmatic examples of power, and the art of persuasion, Gorgias seems to say, is valuable because it makes its possessors powerful.<sup>20</sup> It's a further bragging point that expertise in persuasion is able to get people to do things without the sort of specialized knowledge that, for example, doctors or builders have, and indeed is better at getting people to do what the expertise prescribes than is the expertise itself: for example, without any knowledge of medicine himself, Gorgias is better than his doctor brother at getting patients to follow a prescribed medical regimen (456b).

Against this, Socrates argues that the ability to persuade—to produce conviction without knowledge—is no expertise, for it neither produces a genuine good nor has an account of the means by which it produces what it does (464b–465e). At first, Gorgias' pupil Polus, who takes over the conversation, brushes aside the result that rhetoric isn't an expertise, pointing out that orators are esteemed because they have the greatest power in cities (466a10–b5). Socrates says that they don't seem to him to be esteemed (*oude nomizesthai emoige dokousin*, 466b3), and that if Polus means (*ei . . . legeis*) that power is good for the one who has it, orators don't have great power (466b6–8). Indeed, they have the least power (466b9–10, d7–8) because they don't, so to speak,<sup>21</sup> do what they want (*ouden gar poein hôn Boulontai hôs epos eipein*, 466e1<sup>22</sup>), even

though, like tyrants, they kill, exile, and so on and in general do whatever they think best (466b9–e2). Socrates says Polus himself denies (*ou... phêsin*, 466e4) that having great power is doing what seems best when one isn't doing what one wants, since (*epei*) he says (*ephês*) that power is good for its possessor (466e6–7, cf. 468e1–2). And (as Polus agrees), doing what one thinks best when one doesn't have intelligence (*noun mê echôn*) isn't good for the agent and isn't having great power (466e9–11). So, Socrates concludes: Polus needs to show him that orators are intelligent (*noun echôntas*)<sup>23</sup> and that rhetoric is an expertise (466e13–467a1). Polus heartily agrees (*egôge*, 467a7).

Notice that convincing Polus of this doesn't require Socrates to distinguish doing what seems best from doing what one wants. Yet Socrates persists, 'How do orators or tyrants have great power in cities if Socrates has not been refuted by Polus [and shown] *that they do what they want?*' (467a10). Apparently, Socrates takes 'has intelligence' to be replaceable in the argument by 'does what one wants' and 'does what one wants' to be necessary for 'has power'. But why replace an uncontroversial contrast, between acting with intelligence and acting on uninformed opinion, with an apparently outrageous contrast between doing what one wants and doing what seems best?

The dialectic between Socrates and Polus/Gorgias suggests an answer: the distinction between doing what one wants and doing what seems best allows Socrates to introduce a new conception of power, different from Gorgias' zero-sum, who-has-power-over-whom conception. On Socrates' conception, power has to do with the relationship between the agent and the good the agent wants, and competition with others isn't built into the very notion of power. It is possible to be powerful in Gorgias' sense, and to be able to compel others to do as you say, or to do what seems best to you, but yet not be powerful in Socrates' sense of being able to do what you want or to get the end for the sake of which you are acting. Indeed, the ability to compel others may increase the likelihood of your doing things you do not want to do, and so of acting unwillingly (509e): this is why orators are 'least powerful' (466b) unless they have intelligence. Thus the person Polus thinks is esteemed in cities, the tyrant or orator who does wrong and escapes punishment, does not do what he wants to do, but instead does what he doesn't want to do, and so acts unwillingly—if wrongdoing and escaping punishment are actually bad for him.

To see how all this follows, let us begin with Socrates'—much-maligned—argument for the distinction between doing what one wants and doing what seems best.

- (G1) Whenever someone does something for the sake of something (*ean tis ti prattêi heneka tou*), he doesn't want that which he's doing (*ou touto bouletai, ho prattei*), but that for the sake of which he is doing it (*all' eikeino, hou heneka prattei*) (467d6–e1).

Socrates supports (G1) by examples of actions and the ends for which they are done: one wants health, not the drinking medicine and suffering which one does for the sake of health; one wants wealth, not the sailing and undertaking other risky ventures which one undertakes for the sake of wealth (467c–d).

Commentators sometimes object that these examples are loaded, for sometimes we perform actions for their own sake, and sometimes want both the action we do and the end for the sake of which we do it (we might be the sort to enjoy adventure on the high seas; Polus' tyrant could enjoy killing his enemies as well as wanting all obstacles to power out of the way).<sup>24</sup> But suppose that Socrates means to restrict our attention to instrumental desires. When he talks about doing something for the sake of something else, he excludes doing it for its own sake, and uses the 'loaded' examples to make it clear that this is what he is doing.<sup>25</sup> And even if one and the same action is end and means—say, because the latter is constituted by the former, as dancing is made up of dance-steps—it is possible to inspect the action qua means and qua end ('take steps in time to the music' vs. 'dance') and in this case too to say that it's dancing that one wants and taking steps in time to the music that one does in order to dance.

Restricting Socrates' discussion with Polus to instrumental actions and instrumental desires provides an alternative to supposing that Socrates is assuming psychological eudaemonism (the view that we only want to do what we believe to be the best of the things we can do, and that we either do not have or are not moved to act by any motivations contrary to this desire).<sup>26</sup> One reason to prefer the restrictive alternative is that later in the dialogue, in discussion with Callicles, Socrates says that we have in our soul appetites for such things as food and drink, which he seems to contrast with our desires for the good and our desires to do the best thing we can do (discussed further in Section 3.4.2). It is reasonable for

Socrates to restrict himself to instrumental actions and desires when talking to Polus. When Polus spoke of the power of tyrants and orators he adduced examples of their exiling, expropriating, and killing, presumably thinking of the impunity with which people in such positions can commit actions that most people are held back from by the absence of opportunity or threat of sanctions. But if the tyrant or orator really is an enviable (*zêlôton*, 469e1) person and not just someone fond of exiling, expropriating, and killing for their own sake, Polus must think that these are instrumental actions that help the orator or tyrant achieve his ends. In (G1), by getting Polus to focus on the instrumentality of such actions, Socrates shifts the terms of assessment of the agent's power from what he can do and get away with, as compared to others, to whether he can effect his ends. While some may be more powerful than others on this conception of power too, comparison to others isn't built right into the account of what it is to be powerful.

Socrates continues:

- (G2) There are good things (e.g. wealth, wisdom, health), bad things (their opposites), and intermediate things—those that share sometimes in the good, sometimes in the bad (e.g. sitting, walking, sticks, stones) (467e–468a).
- (G3) When we do one thing for the sake of another, we do intermediate things for the sake of good things (rather than good things for the sake of intermediate things) (468a–b).
- (G4) Therefore (by (G1) and (G3)), we do intermediate things in pursuit of the good (*to agathon ara diôkontes*), for the sake of the good (*tou autou heneka, tou agathou*), thinking that it (viz. the action we do, cf. 468b6) is better (*oiomenoi beltion einai*) (468b1–4).
- (G5) Therefore we don't want to kill, exile, and so on simpliciter (*ouk ara sphattein boulometha oud' ekballein . . . haplôs houtôs*); we want to do these things (*boulometha prattein auta*) only if they are beneficial (*ôphelima*),<sup>27</sup> but if they're harmful (*blabera*) we don't want to do them (*ou boulometha*) (468c2–7).

Although (G1), when I do x for the sake of y, then I want y but don't want to do x, seems to conflict with (G5), when I do x for the sake of

y I want to do x not simply, but on the condition that it is beneficial, and only on that condition,<sup>28</sup> Socrates' thought is quite clear. In (G1) the point is that of the two, end and action that is the means to the end, one prefers the end; in (G5) the point is that the attitude one has to the means is also one of wanting, but it is somehow qualified and somehow conditional on the means really being a means to one's (preferred) end. We might say that (G1) is concerned with wanting (or wanting to have) and (G5) with wanting to do, and Socrates' point about the latter is that although wanting is the attitude we have towards the end, not the actions we undertake to bring it about, wanting the end is the basis for a derivative desire, to do actions which are means to the end, and this derivative desire is conditional on the actual relationship between end and actions undertaken for its sake.<sup>29</sup>

However, this still leaves (G5) open to different interpretations. On one influential interpretation, (G5) says:

(G5\*) I want to do x only on the condition that doing x is a means to y (y being a good I want); if doing x is not a means to y, I do not want to do x.<sup>30</sup>

(G5\*) distinguishes the object of my desire (to do x) from the condition that must be met (viz. that x is a means to y) for me to have a desire capable of fulfilment or frustration; in the absence of that condition, we can say that my desire is 'cancelled'.<sup>31</sup> 'Cancelled' on this analysis would correspond to Socrates' 'doesn't want to do' in the argument above. According to (G5\*), as far as wanting goes, when doing x is in fact a means to y, I have the same (unqualified, unconditional) pro-attitude towards doing x that I have towards y; when doing x is not a means to y, I lack this pro-attitude.

However, Socrates seems to want to mark a difference between the desirer's attitude towards his end (y) in (G1) and towards his means (doing x) in (G1) *even if doing x is in fact a means to y*. So I propose an alternative interpretation of (G5):

(G5\*\*) I want to do x only on the condition that doing x is a means to y (y being a good I want), but even if this condition obtains, my wanting to do x is different from and weaker than my wanting y; in the condition that doing x is not a means to y/undermines y's coming-about, I don't want to do x.



We may usefully compare wanting to do in (G5\*\*) with selection with reservation, the Stoics' prescribed attitude towards the indifferents we aim at in our actions. A Stoic might say: 'I select health as according to nature, but with the recognition that it might turn out not to be according to nature for me to have it (for I lack knowledge of Zeus' plan for the world)—with the result that if I don't have health despite my efforts, I will not suffer a negative emotion.<sup>32</sup> But even if it is according to nature for me to be healthy, I don't select it unconditionally—for the only thing I go for unconditionally is virtue or being in agreement with nature. And obtaining the health that I selected doesn't make me happy, although understanding the place of my selecting and obtaining health in the whole system of nature would.'

An advantage of (G5\*\*) over (G5\*) is that retaining the information that my pro-attitude to doing *x* is derived from its being instrumental to *y* may explain other things I do in relation to *x* in addition to trying to do it. Consider Polus' tyrant: the statement, 'he wants to exile so-and-so' seems on the face of it to conflict with his saying things like, 'I don't really want to exile so-and-so' and 'I don't want to face so-and-so', and his losing interest in exiling so-and-so if so-and-so drops out of politics. But the statement, 'he's exiling so-and-so because he wants unchallenged power' explains more actions and removes the apparent conflict. An action that we undertake as a means to some end may have many features to which we are averse and towards which we express our aversion, but the fact that the action in question is a means to an end we want puts such expressions in their proper context.

The relationship between wanting *y* and wanting to do *x* when *x* is a means to *y* depends on a principle of practical rationality: if you will the end, you will (the/some/best) means, or more circumspectly, if you will the end, you don't will (you anti-will), for the sake of that end, things that in fact hinder its realization. It is presumably on the strength of some such principle that people say about others, 'she wants to do such-and-such', or 'she doesn't want to do such-and-such' in the absence of explicit avowals of wanting or not wanting to do such-and-such. More rarely we say such things contrary to explicit avowals, but the reason this is rarer is that contrary avowals call into question our grip on the desirer's end. Socrates in fact only uses the more circumspect negative principle, to deattribute instrumental desires on the grounds that they prevent, rather than bring about, the end for the sake of which the agent is acting.

It may be objected to the restriction of this discussion to instrumental actions that later in the dialogue, Socrates says to Callicles, 'Remember, it seemed to Polus and me that we should do everything we do for the sake of good things' (*heneka gar pou tôn agathôn hapanta hêmin edoxen prakteon einai*, 499e); 'everything' is not restricted here and yet Socrates is recalling an agreement with Polus. However, this statement, about what *should* be done (cf. *dein panta t'alla prattesthai*, 500a1), cannot be referring to our (G3), according to which we *in fact* do instrumental actions for the sake of good things. There are two differences between (G3) and Socrates' claim at 499e: first, (G3) makes a claim about instrumental actions, whereas 499e makes a claim about all actions; second, (G3) is a descriptive claim whereas 499e is a normative claim. Socrates and Polus agree to the normative claim only at 480a, after Socrates has established that injustice is a sickness in the soul and likened paying the justly imposed penalty to undergoing medical treatment: a person *should* guard against doing injustice, and if he or anyone he cares about does injustice, he *should* willingly pay the penalty, and oratory is useful only to accuse doers of injustice so that they may pay a just penalty and thereby be rid of their injustice.

I have left for last the biggest objection commentators raise for Socrates' argument, which is that the sense of 'good' shifts, from 'believed by the agent to be good' in (G2) and (G3), to 'actually good' by (G5), where 'beneficial' must mean 'productive of what is really good'.<sup>33</sup>

However, the argument (G1)–(G5) doesn't require a such a shift in sense and 'good' can be read from beginning to end either as 'believed to be good' (subjectively) or as 'really good' (objectively). On a subjective reading, 'beneficial' in (G5) would not mean 'productive of the real good', but rather, 'really productive of the agent's end, where the end is believed by the agent to be good', and the point would be: we want to do the actions we do as means only if they are *in fact* productive of the ends (which we believe to be good) for the sake of which we are doing them. On this reading, Socrates' point would be, in contemporary jargon, that instrumental desires have the same direction of fit as beliefs: they are to fit the world, not the other way round.

On an objective reading, which is more plausible, (G2) and (G3) would already assume that the goods we pursue, for the sake of which we are acting, are real goods. It might be objected that Socrates includes in the class of good things health and wealth, both of which he elsewhere

classifies as beneficial (good for us) only if used with wisdom (*Euthydemus* 280c–281e, *Meno* 87d–89a).<sup>34</sup> It is probably correct that if we were to ask whether health and wealth are unconditionally good, or good only on the condition that they are controlled by wisdom, Socrates would have to class them with the intermediates, since they sometimes partake of good, sometimes of bad, sometimes of neither, just like sitting, walking, sticks, and stones. But (G2) in our passage is not classifying things by whether they are good by nature/unconditionally or only conditionally; it is classifying them by whether they are final or instrumental goods—the intermediates being things like actions that can be means, and objects that can be tools, and in that way ‘partake of good’. It’s a separate question in what combinations we need to have the final goods for them to be beneficial/good for us/such as to make us happy. Further, while wealth is, considered absolutely, an instrumental good, relative to the sea voyage (Socrates’ example), it is a final good.

Finally, Polus should not balk at objective readings of (G2) and (G3), for his praise of rhetoric only succeeds if rhetoric secures real, not merely apparent, goods for its practitioner. In accepting ‘good’ in the sense ‘really good’, Polus would be in step with interlocutors throughout the dialogues who assume that when it comes to our own advantage or good or happiness, it is the real thing that we want. For example, when Thrasymachus’ definition of (legal) justice, ‘the advantage of the rulers’, is faced with the possibility of rulers legislating contrary to their own advantage, he refuses to save his position by modifying his definition to ‘justice is what the rulers believe is their advantage’, and instead claims that the ruler insofar as he possesses the *expertise* of ruling never legislates to his disadvantage (*Republic* 340a–c). In the *Theaetetus*, Protagoras’ relativism—things are for each perceiver as they seem to that perceiver—makes an exception of goodness or advantage, about which Protagoras maintains that one person is wiser than another, and about which neither law nor the community can determine how things are by how things seem (167a–d, 171e–172b, 177d–178a). This widespread agreement about our attitude towards the good is reflected in Socrates’ statements, in the *Republic* (505d–e) and *Philebus* (20d), that every soul pursues the good in its actions. This is why in the *Gorgias* Callicles responds to Socrates’ speech about the badness for the agent of doing injustice and the melioration of being punished, ‘Tell me, Socrates, are we to take you as being in earnest now, or joking? For if you are in

earnest, and these things you're saying are really true, won't this human life of ours be turned upside down, and won't everything we do evidently be the opposite of what we should do?' (481c).<sup>35</sup>

Returning to our argument (G1)–(G5) from the *Gorgias*, it's worth underlining that while (G3) says that what seems best to the agent explains his actions (Q: why was he walking? A: he thought walking was best), this explanation is very local. That I seek what's really good is required to explain my doing what seems best.<sup>36</sup>

This reconstruction of (G1)–(G5) as explicating an alternative to Gorgias' conception of power agrees with previous scholarly accounts in finding in 'wanting' a notion of rational desire. However, this reconstruction emphasizes the truth-trackingness of instrumental desire.<sup>37</sup> It will be useful to say briefly why I favour this over alternative accounts of wanting. Some commentators say that wanting (*boulêsis*) is the desire of my true—that is, rational—self.<sup>38</sup> But while Plato may hypothesize such an entity in the *Republic*, he does not in the *Gorgias*, and in any case, identifying such an entity doesn't explain what is rational about its desire. Saying that what I want is what I would want if I were fully informed<sup>39</sup> does answer this question, but then calls for an explanation of why Socrates says, 'they don't do what they *want*' if he means 'they don't do what they *would want, if...*'. Or to make the point philosophically rather than textually: it seems that coming to know what I do not now know would change at least some of my desires, so how is what I would want with changed desires normative for me now? Socrates' statement, made in terms of what we want (in the present), assumes that at least one of our present desires is immune to being changed by what we come to know, and can only be fulfilled (or not); Socrates does not need to argue for its normativity over us now because we already have that desire. Restricting the attitude of wanting to the knower, so that I can only want something when I know it is good,<sup>40</sup> runs afoul of Socrates' assertion that if we are actually to avoid wrongdoing, we need not only to want it; we need knowledge and power in addition (509d–10a). Appeal to *Charmides* 167e, where Socrates says that *epithumia* (appetite) is for pleasure and *boulêsis* (wanting) for good, doesn't by itself say what makes *boulêsis* rational, or even tell us whether on this distinction the good that is the object of *boulêsis* is real or apparent. Finally, even if we think what makes wanting special is its being keyed to what is really good, which

makes sense of our desiring and trying to get what we believe good, searching beyond appearances/opinions for what is really good, and sometimes being discontented even when we get the thing we previously thought good,<sup>41</sup> this doesn't make any use of the distinctive attitude towards instrumental actions in Socrates' actual argument (G1)–(G5) above.

In the *Nicomachean Ethics* Aristotle stipulates that the name for our attitude towards the actions we undertake as means to our ends is *prohairesis* (choice, decision), contrasting it with *boulêsis*, which is for the end (which according to Aristotle is the apparent good (III.4, 1113a15–22)<sup>42</sup>). His reasons for introducing a separate term are that whereas we choose what promotes our end, is possible, and is up to us to do, wanting is for an end and we may want something impossible or want someone else to do something—but we cannot choose these (III.2 111b7–30). The last of Aristotle's claims about wanting is corroborated by the fact that *boulê* (cognate with Latin *volo* and English 'will') is used in Homer of a god's will or plan, often executed by the agency of another (*Iliad* 1.5, *Odyssey* 3.232) and also of a man's [wise] counsel to someone else about what he should do (*Iliad* 1.258, 2.202, 24.652). Despite this linguistic legislation, Aristotle's discussion shows its roots in Plato's *Gorgias*: *boulêsis* is *more* or *rather* for the end, he says, contrasting our wanting health with our choosing the things through which we are healthy (1111b26). While Aristotle's use of 'choice' (*prohairesis*) for our attitude towards our own purposive actions, distinct from 'wish' (*boulêsis*) for our attitude towards ends, avoids the confusions we saw above arising from the *Gorgias*' use of 'wanting' (*boulesthai*) for both, in Aristotle the dependence of wanting to do on wanting the end is no longer marked linguistically. Instead, Aristotle's characterization of thought as deliberative desire (*bouleutikê orexis*, 1113a11) emphasizes the rationality of the deliberative process (*bouleusis*) by which wish (*boulêsis*) gives rise to choice (*prohairesis*).

Socrates' claim that one can only want (*boulesthai*) what is really good as an end or what is really productive of this good suggests a two-fold reply to the objection so many readers have to this argument, well expressed in Dodds' complaint, 'Whatever its theoretical justification . . . what I "really" want to do turns out to be a polite paraphrase for what you think I ought to want to do.'<sup>43</sup> First, it isn't just a question of two

different opinions about what one ought to want to do, because there are facts about which actions are means to a given end that bear on whether or not one can have an instrumental desire towards certain actions. Second, there are facts aside from how an agent represents an object to herself that bear on whether it is true of her that she desires it. G. E. M. Anscombe famously says that a primitive sign of wanting is trying to get,<sup>44</sup> but it is not the only sign: a second primitive sign of wanting is of being fulfilled and at rest once one has gotten what one was trying to get (cf. Plotinus *Enneads* I.4.6.19).

Let's return to the last steps of Socrates' argument for the distinction between doing what one wants and doing what one thinks best, the distinction he needs in order to be able to say that sometimes agents don't want to do the actions they do, and so act unwillingly, apparently irrespectively of their avowals:

- (G6) If the orator or tyrant kills or exiles, for example, thinking it's better but it turns out to be worse, he is still doing what he thinks best (468d1–5).
- (G7) But (by (G5)) if killing, exiling, and so on turns out to be (i.e. results in) a bad, he is not doing what he wants to do (468d5–7).
- (G8) If being powerful is good, then the person who can kill, exile, and so on when it seems best to him is not thereby powerful (468d7–e3).
- (G9) It's possible that a person who does what seems best to him in the city (in (G6)) neither has great power (by (G8)) nor does what he wants to do (by (G7)) (468e3–5).

Socrates harks back to this argument when he claims, later in the *Gorgias*, that no one commits injustice wanting to (*boulomenon*),<sup>45</sup> but all who do injustice do so unwillingly (*akontes*) (509e). That is to say, the agent of injustice, even if he does what seems best to him, not only isn't doing what he wants, but also doesn't want to do what he does. This is because:

- (G10) People do unjust things in order to obtain some good.

(G10) is implied in Polus' claim that tyrants like Archelaus are happy (470d): he is not saying that tyrants happen to be both happy and unjust;

rather, he claims that it is by means of doing unjust actions and getting away with them that tyrants obtain good things and are happy. Polus puts unjust actions in the class of things that are sometimes beneficial, sometimes harmful—like the sticks and stones, walking, sitting, and so on of (G2) and (G3). Socrates means to put them in the class of things that are always harmful.

- (G11) Doing what's unjust is bad (indeed the worst thing) for the agent of the injustice (469b), worse than suffering injustice is for the sufferer; not paying the penalty for one's injustice is worse for the unjust person than is paying the penalty (474b).<sup>46</sup>
  - (a) For doing what's unjust is more shameful than suffering injustice, but not more painful, yet something can only be more shameful by being more painful or worse<sup>47</sup> (474d–475d).
  - (b) Punishing, that is, imposing a penalty, justly is admirable, and consequently so is being subject to the just penalty; since paying a penalty is painful, the admirability of paying a just penalty must be due to its goodness—that is, its getting rid of the bad, the injustice, in the penalty-payer's soul (476b–479d).
- (G12) Therefore, the agent who does unjust actions (for the sake of good things) is not doing what she wants (which would have to be something beneficial), but is doing something she does not want (since it in fact makes her worse).
- (G13) If you are not doing what you want but doing what you do not want, then you are acting unwillingly.

According to this argument, an action is unwilling if it is an action that the agent does not want to do, and an agent does not want to do actions that hinder the end for the sake of which she is doing them. So, because unjust actions hinder the happiness for the sake of which the agent does them, no one wants to do unjust actions, and if an agent does unjust actions, she does them not wanting to, but unwillingly.

The plausibility of this conclusion depends on the means–ends structure into which Socrates has fit unjust actions. For while it's not obvious which of two conflicting desires (for happiness, to act unjustly) to privilege as a person's real desire, putting those desires into a means–

ends hierarchy privileges the end. If my end is happiness and my supposed means to happiness is unjust actions, but in fact unjust actions are a sure-fire means to unhappiness, it does seem reasonable to say that I am doing those unjust actions contrary to my desire (for happiness, and so to do things which bring about my happiness and not to do things that detract from it), and so unwillingly.

In sum, in the *Gorgias*, Socrates argues that doing injustice is unwilling as follows: unjust actions are done for the sake of happiness; doing injustice makes the agent unhappy; an action undertaken for the sake of some end that it actually undermines is unwilling. On this account the argument that the action is unwilling does not appeal to the agent's being ignorant, although ignorance can explain why someone would undertake an action for the sake of an end when it undermines that end. Further, the ignorance that explains is not merely ignorance of the fact that injustice is bad for the agent (which could be rectified by acquiring the true belief that it is bad<sup>48</sup>), but the more difficult-to-rectify ignorance of which actions actually conduce to the ends one is trying to realize.

### 3.4.1 Attribution and Deattribution of Psychological Attitudes

Socrates' claim that the tyrant 'doesn't want' the actually bad action he does even though it seems best to him is paralleled by his saying, in the *Gorgias*, that people 'don't believe' or 'don't say' things they avow. For example, against Polus' protests to the contrary, Socrates claims: (a) Polus doesn't say (*phêsin*) that doing what seems best is having great power (even though Polus does say it) (466d–e); (b) it seems to Polus (*phainetai soi*) that one doesn't want what one does unless it is beneficial (470a); and (c) Polus and Socrates and every other human being consider (*hêgêsthai*) doing injustice to be worse than suffering it (even though Polus asserts that suffering injustice is worse) (474b).

For at least (a) and (b), Socrates is explicit that the grounds for his attributions are Polus' other assertions: it is 'since' Polus says (*epei*, 466e6; *eiper* . . . *kata tèn sên homologian*, 468e1–2) that having power is a good that he believes one doesn't want to do things if they are not beneficial. Similarly, we might suppose that Socrates' basis for (c) is that Polus and Socrates have agreed that doing injustice is more shameful/less admirable than suffering it, and that if it is more shameful/less admirable, then it must be either more painful/less pleasant or worse/less good, but that doing justice is clearly less painful/more pleasant than suffering



injustice. So Socrates' 'you believe' and 'you say' would be a claim about Polus' commitments, given his avowals.

However, Socrates' attribution (c) precedes Polus' assertion that it is more shameful to do than to suffer injustice; Socrates attributes the belief to all human beings; finally, Socrates says quite generally that orators aren't esteemed in cities (*ou nomizesthai*, 466b3). Does he know what everyone believes before asking them anything?

Gregory Vlastos argues that Socrates' practice of elenchus is based on the assumption: 'whoever has a false moral belief will always have at the same time true beliefs entailing the negation of that false belief'. Vlastos thinks that Socrates' assumption is founded on his long experience of cross-examining people, and that Plato, noticing what a strong assumption this is, puts it on display in the *Gorgias* and supplements it in the *Meno* with the account of recollection, according to which we are all in possession of the truth and need only recollect it, recollection being a process enabled by elenchus.<sup>49</sup> Might we suppose that experience also informs Socrates' claims about what people want, and that Plato thinks recollection, or some conative analogue, accounts for our wanting the good? If so, Socrates would be on firmer ground with attributing and deattributing desires than he is with beliefs, since among beliefs he can only identify a conflict and then rely on experience to tell him which belief is more deeply held, whereas many conative attitudes belong to a means-ends hierarchy which enables us to easily identify which want is privileged.

But even if Socrates' claims about what all human beings believe are rooted in long experience of refuting people and/or a theory of recollection waiting in the wings, it's also worth noting that it's not only Socrates, but also Polus, who claims to know what people believe. Recall that one of Gorgias' arguments for the great power of rhetoric in the *Encomium* is that most of us rely on belief (*doxa*), and persuasive speech is able to produce any belief it likes in its audience. Add to this that in the *Gorgias* Polus' mode of refutation is to show his interlocutor's nonconformity with what other people believe. So certainly he, but presumably any orator who needs to change his audience's beliefs by persuasion, needs to know what people believe. Polus says that in cities, people esteem orators (466a). Polus also claims to know what people really believe despite what they say: Gorgias is too ashamed to say what he really thinks about orators teaching justice (461b-c); Socrates agrees with

Polus that injustice is a means to happiness but just won't say it (471d). Polus doesn't 'know' all this by observation alone, but on the basis of a theory of human nature which says what people really want—wealth, power, and so on—the pursuit of which is hindered by justice. It follows that only a fool or a liar would say that doing justice is better than injustice. By contrast, Socrates says he refutes his interlocutor using only the interlocutor's attestations (471e–472c, 473d–474b, 475e–476a), showing that whereas Socrates doesn't agree with the majority, the interlocutor doesn't even agree with himself, and so doesn't even know what he believes (cf. 482b–c).<sup>50</sup> Whereas Polus' attributions of beliefs and desires are based on a theory of human nature that takes for granted what the goods are that human nature seeks, Socrates' arguments show the revolutionary potential of this theory depending on what actually, upon examination, turns out to be good.

Socrates' deattribution of false beliefs about what is good is bound to conflict with his attribution of false beliefs about the good to explain actions (using the *dokei einai* locution to contrast with wanting).<sup>51</sup> That is to say, the parallelism between believing and wanting I've just drawn out would have Socrates say, 'the tyrant doesn't believe injustice is better than justice', but my analysis of wanting would have Socrates say 'the tyrant performed these unjust acts not because he wanted to but because he believes injustice is better than justice.' Belief-attribution is governed by different standards in theoretical and practical contexts: in the former, we attribute belief on the basis of consistency of the proposition in question with the subject's other beliefs, the subject's commitment to the entailments of the proposition in question, and so on; in the latter, we may say 'X seemed best to him' just on the strength of his doing X for the sake of Y without attention to the theoretical standards for belief-attribution. Perhaps it makes a difference that the belief attributed in the theoretical context is treated as dispositional (i.e. involving the disposition to make inferences based on the proposition, accept its entailments, etc.) whereas the belief attributed in the practical context may be occurrent and one-off.

Socrates comments on the parallelism between wanting and believing in a speculative etymology in the *Cratylus*, when he says:

'Belief' (*doxa*) is named after the pursuit (*diôxei*), to which the soul pursuing knowledge of how things are is carried, or after the shooting (*bolêi*) of a bow

(*toxou*) . . . just as indeed deliberation (*boulê*) is somehow shooting and ‘wanting’ (*boulesthai*) and ‘deliberating’ (*bouleuesthai*) signify aiming. All these seem to follow belief insofar as they are likenesses of shooting. (420b–c)<sup>52</sup>

‘Shooting’, here, is Plato’s metaphor for what I have been calling truth-tracking in the case of belief and instrumental desire.

### 3.4.2 *Appetitive Desires*

The truth-tracking character of wanting challenges the conception of the soul in Gorgias’ *Encomium* (echoed by the beginning of Plato’s *Gorgias*) as deprived of memory, foresight, and understanding, and therefore passive and impressionable before the persuasive power of speech. For in wanting, the soul has a connection to the truth which can be a source of resistance to persuasive speech. Still, in the *Gorgias* Socrates seems to concede that there is something in the soul that fits Gorgias’ characterization of the soul as a whole. In his discussion with Callicles Socrates describes a class of desires, which we may call ‘appetites’ (*epithumiai*), that he says need to be disciplined: these are located in a part of the soul that is persuadable, a part that like a jar may be ‘filled’ and then may stay filled or leak; Callicles adds that filling brings pleasure (491d–494a).

The appetites are not instrumental desires because they are for filling rather than for doing anything in particular to bring the filling about. In this respect the appetites are like the wanting for the end that was the source of our instrumental wants or wanting to do actions. It belongs to something else in the soul to do the filling, and Socrates says that in the soul of a fool this agent of filling is like a sieve (imagine this as a ladle, which may or may not be holey). Since the appetites are not instrumental desires, it does not seem possible to use the logic of instrumental rationality to deattribute them from the agent. Indeed, Socrates’ emphasis on their persuadability seems to pointedly contrast them with truth-tracking instrumental desires. Appetites’ persuadability and the pleasure that results from their being filled suggests that they are for what appears good. But Socrates does not explain how the two elements, jar and ladle, or appetite and agent of fulfilment, coordinate.

It is standard in the scholarship to describe the difference between appetites and wanting as a difference between good-independent and good-dependent desires. Characterizing the appetites as good-independent desires might follow from their origin in bodily conditions

and anticipate the *Republic* (on a reading Chapter 4 will dispute). But it does not explain how they are particularly persuadable. We can, however, understand their persuadability if we pay attention to the contrast, running through the *Gorgias*, between expertise (*technê*) and knack (*empeiria*), the former aiming at the good and based on causal knowledge, the latter aiming at pleasure and based on experience (463a–465e, 500a–501c). At least some of the knacks flatter the appetites (499d, 503c), and they do so by producing something that appears to be the good that the expertise they mimic produces: an apparent state of bodily fitness, for example (464a), or apparently good food (464d).<sup>53</sup> If appetites are directed at an apparent good, then perhaps they can give rise to actions in the way that wanting does.<sup>54</sup> Still, it does seem important that the source of at least some appetites is bodily (after all, the appetites seem to arise in the soul when it is ‘entombed’ in a body), and that the thing to do with appetites is not simply fulfil them but also discipline them.<sup>55</sup>

Given that in the *Gorgias*, Socrates recommends disciplining the appetites so that they do not need much filling, against Callicles’ claim that the best way to live is to let one’s appetites grow large and be brave and clever enough to fulfil them, it is surprising that in the *Meno*, Socrates objects to Meno’s proposed definition of virtue, ‘desiring fine things, to be able to procure them’ (*epithumounta tôn kalôn dunaton einai porizesthai*, 77b4–5), that desiring good things is common to everyone (*pantes . . . tôn agathôn epithumein*, 77c1–2; *to men boulesthai pasi huparchei*, 78b5) and so isn’t a part of virtue (78a).

Socrates’ response to Meno’s definition begins by fixing the terms: ‘Do you mean the person who desires fine things is a desirer (*epithumêtên*) of good things?’ (77b6–7). The question fixes Meno’s meaning in two ways: most obviously, it substitutes ‘good things’ for ‘fine things’; less obviously, it picks up on the fact that Meno has defined virtue in terms of the virtuous person, suggesting that what he says about virtue is true of all and only virtuous people. For Socrates has asked for a definition that identifies the form which picks out all and only cases of virtue as virtue (72c). We can see the less obvious point in the way Socrates continues: ‘[Do you mean] that there are some who desire bad things, and others who desire good things? Don’t they all, good man, seem to you to desire good things?’ (77b7–c2). Meno thinks people are distinguished into virtuous and vicious both by whether they desire good things and by

whether they are able to get them. As Socrates questions him about the desire criterion, Meno divides people up first into (1) those who desire good things and (2) those who desire bad things; then, among those who desire bad things, into (2a) those who desire bad things knowing that they are bad and (2b) those who desire bad things believing that they are good; finally, he divides those who desire bad things knowing that they are bad into (2ai) those who believe that having the bad things will benefit them, and (2aii) those who know that having the bad things will harm them. The people in (2ai) turn out not to know that the bad things they desire are bad, and so must belong to (2b), those who desire bad things believing that they are good. I will say more about (2b) later in this section.

As for (2aii), Socrates argues that there are no such people. This is because people who desire bad things knowing they are bad and knowing that bad things harm their possessors would know that those who are harmed are wretched and unhappy to the extent that they are harmed (cf. *Apology* 25c–26a, discussed in Section 3.3.2). However,

s: Is there anyone who wants (*bouletai*) to be wretched and unhappy?

m: Doesn't seem so to me, Socrates.

s: Then, Meno, no one [wants] bad things, unless he wants (*bouletai*) to be such. For what else is it to be wretched than to desire (*epithumein*) bad things and get them? (78a4–9)

The conclusion, then, is that Meno's definition of virtue—now rephrased by Socrates as 'both wanting (*boulesthai*) good things and being able [to get them]' (78b3)—has a component that does not make one person better than another, for 'the wanting belongs to all' (78b5). Since wanting good things belongs equally to virtuous and vicious people, Socrates and Meno are now left with the idea that since one person is better than another in ability (*kata to dunasthai*), this must be what virtue is (a definition with which Socrates will also find fault).

Meno could legitimately point out in reply, first, that his initial definition of virtue said nothing about wanting (*boulesthai*) but only spoke of desiring (*epithumein*), and second, that Socrates' own definition of unhappiness, desiring (*epithumein*), and getting bad things, suggests that what we desire (*epithumein*) is relevant to our virtue or vice after all (assuming that virtue and vice bear on happiness and unhappiness). However, the definition of vice may include the desire for bad things even if the definition of virtue does not include the desire for good things, if

both the vicious and the virtuous desire good things but the virtuous do not desire bad things. And the point about the difference between wanting and desiring is fair, but Socrates' target is Meno's division of people into virtuous and vicious by what they desire without any acknowledgment that at a fundamental level they want the same thing, diverging because of what they rightly or wrongly believe to be good.<sup>56</sup>

But, it may be objected, doesn't Socrates deny that anyone, virtuous or vicious, desires bad things? Let's look a little more closely at what he says about the people Meno classifies under (2b), people who desire bad things believing that they are good:

Then it's clear that on the one hand, *they* don't desire the bad things (*dêlon hoti houtoi men ou tôn kakôn epithumousin*), the ones who are ignorant about them (*hoi agnountes auta*); rather, [they desire] the things they think are good (*ekeinôn ha ôinto agatha einai*), but these things are in fact (*estin de tauta ge kaka*) bad, with the result that the people who are ignorant about them (*hoi agnountes auta*), and think they are good, it's clear that they desire good things (*dêlon hoti tôn agathôn epithumousin*). (77d7–e3, my emphasis)

Having shown that the people (in 2ai) who (supposedly) desire bad things knowing they are bad, but not knowing that bad things harm their possessors, are people who desire these things thinking they are good (2b), Socrates is now showing that these people desire good things—in other words that they belong to Meno's group (1) above, people who desire good things.

In this much-commented-on passage, Socrates says three things about the people in (2b), who Meno thinks desire actually bad things thinking they are good: (i) they are ignorant about these things (presumably of their badness); (ii) they don't desire them; (iii) they desire the things they believe are good, and so they desire good things. How are these points connected?

On one influential interpretation of these lines, Socrates' point is: when people desire bad things they believe them to be good; therefore, they desire things they believe to be good; therefore, they desire apparently good things.<sup>57</sup> But it would not serve Socrates (or Plato) to infer 'they desire apparently good things' from 'they desire what they believe are good things', for then Socrates would after all be conceding to Meno two of the kinds of people from his original classification, (1) those who desire really good things and (2b) those who desire apparently good things—in which case, it would be misleading for him to say that the desiring is common to all.<sup>58</sup> Further, Socrates' emphatic language

suggests that he is making a rather bigger point than substituting ‘apparently good things’ for ‘things believed to be good’. Finally, the distinction he has just made between things that are actually good and things that are only believed to be good would be undermined by calling the things believed to be good, simply, ‘good things’.

On another influential interpretation, Socrates believes that it is impossible to desire anything unless it is actually good.<sup>59</sup> But how is this to be reconciled with Socrates’ account of unhappiness, desiring bad things and getting them (78a9)? Is no one unhappy? Or why would Socrates be proposing a faulty definition of unhappiness?

Agnes Callard argues that Socrates thinks there are two necessary conditions for desiring something: that it actually be good and that the desirer believe it is good, and that while in the *Gorgias* he emphasizes the former, in this passage in the *Meno* he emphasizes the latter.<sup>60</sup> *Meno* 77d–e distinguishes two groups of things, namely the bad things (*auta/tauta*) which are not desired and the things thought good (*ekeinôn*), which are. The reason the bad things are not desired is that the people who are said to desire them don’t know them, but to desire them they must have some cognitive contact with them.

We should grant Callard’s point about the necessity of cognitive contact for desiring, for it is obviously important to Socrates’ overall eliminative argument for ‘desiring is common to all’ that the people in group (2b) believe that the things they desire are good. But from this, Socrates reasons that they desire really good things. So (iii) makes the plausible psychological point that the description under which you pursue an object determines what the object of your desire is. For example, if I pick up and drink a glass of clear liquid believing that it’s water, then even if that clear liquid is vodka, it’s water—real water—that I desire. Similarly, if you pursue something believing it is good, or purple, you must desire the real good, or the real purple thing, you think it is (there’s nothing special about the guise of the *good* here). This is what the reasoning of (i)–(ii) must be. In the water–vodka case it is not accurate to say of me, ‘she desires that vodka’; better to say, ‘she desires that vodka believing that it’s water’ or even just, ‘she desires water’. If it is misleading to say of someone who pursues *x* in the false belief that it is *y*, ‘she desires *x*’, then it is misleading to say of someone who pursues bad things in the false belief that they are good, ‘she desires bad things’.<sup>61</sup>

What should the definition of virtue say about desire? From the *Gorgias* we learn that to avoid doing injustice as well as to avoid suffering injustice it isn't sufficient to want (*boulesthai*) good things; one also needs some power or expertise (509e–510a); from the *Meno* we learn that to want (*boulesthai*) good things belongs to human nature, and not just to virtue. The *Gorgias* tells us that virtue involves the control of our appetitive desires (*epithumiai*), which takes some expertise, and the *Meno* brings out a conflict between the thought that appetitive desire for something bad seems to require ignorance of its badness, and the thought that to desire something you have to know it. The last piece in the puzzle comes from the *Hippias Minor*.

### 3.4.3 *Hippias Minor on How It Could Be the Good Person Who Does Wrong Willingly*

The *Gorgias*' conception of wanting to do actions as means sheds new light on the puzzling conclusion of the *Hippias Minor* that the good person, that is, the person with a wise and powerful soul, would be the person who errs and does shameful and unjust things willingly, or when he wants to (366e–367a)—if there is a such a person (376a–b). Many scholars take this conditional conclusion to be contrary to fact,<sup>62</sup> contrary in particular to the Socratic doctrinal fact that no one does wrong willingly. But Socrates describes himself as of two minds about the supposed doctrinal fact (376c–d, cf. 372d–e), and since he says that the conclusion seems to him to follow from the previous argument (372e3–5, 376c1), we should at least try to see how it is supposed to follow. I argue below that if we conceive of virtue as the power to do what one wants without fixing that what one wants is the good, so that what one wants to do must be whatever conduces to that good, then it does follow that the good person would do wrong (as he would do anything) willingly rather than unwillingly.

Socrates summarizes the argument in the last page of the dialogue leading up to the conclusion (375d–376c): if justice is knowledge or power or both (cf. *Gorgias* 509d–e, *Meno* 78e), then the more just soul will be the more knowledgeable and more powerful, and being more powerful, will be the one to do unjust and base and bad things willingly rather than unwillingly. To contextualize: the question has been whether, where the doing of bad things is concerned, the better/good person is the one who



does them willingly or unwillingly, with Hippias championing ‘unwillingly’, since he thinks unwilling bad actions, like Achilles’ false speeches, don’t tell against the virtue of their agent (presumably because they should be excused, as discussed in Section 3.3). According to Hippias, Achilles is the best of the Achaians and certainly better than the shifty and lying Odysseus, and his falsehoods, unlike Odysseus’, are unwilling.

Socrates first reasons: the power in question, by which a person is able to lie when he wants to, and not unwittingly tell the truth, is knowledge of the subject matter about which he lies, and that is also the power by which the truth-teller tells the truth when he wants to. So the liar and truth-teller aren’t opposite kinds, but the same kind; that is, the wise person (366c–369a). Although this may seem to indicate that Socrates identifies people by their powers, it might indicate only that *if* the liar is identified by his power, then he can’t be distinguished from the truth-teller.<sup>63</sup>

Next, Socrates argues for the superiority of the willing wrongdoer across a range of cases (373a–376c). First, in physical activities the one who does bad things willingly is superior (373d–374b). Consider the example of running a race:

- (H1) The one who runs well is good (373d2).
- (H2) The runner who runs quickly runs well; the runner who runs slowly runs badly (373d3).
- (H3) In a race, quickness is good; slowness is bad (373d4–5).
- (H4) Of two runners, the one who runs slowly willingly is better than the one who runs slowly unwillingly (373d5–7).
- (H5) Running is doing/effecting something (*poiein/ergazesthai ti*) (373d7–e1).
- (H6) The one who runs badly effects something bad and shameful in the race (373e1–2).
- (H7) The one who runs slowly runs badly (373e3).
- (H8) Then the good runner effects the bad and shameful willingly; the bad runner effects the bad and shameful unwillingly (373e4–5).<sup>64</sup>

(H1)–(H3) set up the evaluative criteria for running a race: what running well or badly are, what a good and bad runner are. (H4) states the superiority of the willingly slow runner to the unwillingly slow runner, presumably on the grounds that running slowly unwillingly would be due to some weakness in or constraint on the runner. We imagine the

one who runs slowly willingly as able to run quickly if he wants to, but (for some reason) not doing so on this occasion; we contrast him with the one who runs slowly because his weak bodily constitution doesn't allow him to run any faster; we may also contrast him with the one who runs quickly unwillingly—out of fear of a punishing trainer, perhaps. The willing runner, who runs quickly or slowly when he wants to, seems to be free from constraints. (H5)–(H7) spell out that running slowly is bad by the criteria (H1)–(H3), and (H8) uses (H4) to say that nevertheless the good runner will be the one to run slowly willingly. Unless we introduce evaluative criteria that come from whatever the agent might run *for* (as opposed to coming only from running itself), or at any rate from some wider context of evaluation, this conclusion seems irresistible. Why shouldn't it be the same for the willing as compared with the unwilling doer of injustice?

With the next group of cases, of bodily conditions, Socrates argues that it is better to be graceless, have legs that limp, a voice that sings out of tune, eyes that see poorly, and so on, willingly rather than unwillingly (374b–e). For being graceless willingly is on the side of virtue (*pros aretês*), but being graceless unwillingly on the side of vice (*pros ponêrias*, 373b8–c1). In these cases the 'willingly' would seem to attach to the possessor of gracelessness or poor-sightedness, for example. But a gracelessness I can turn on and off at will doesn't seem to be a bodily condition, and it's difficult to imagine how I can see poorly at will (by covering my eyes? crossing my eyes?)—in any case, if I can see poorly at will I don't have the condition, poor-sightedness, which Socrates says is a bad condition of eyes (*amblyôpia . . . ponêria ophthalmôn*, 374d2–3). If the argument is parallel to the argument about physical activities exemplified above, Socrates' point should be that it's better to have the bad things willingly ((4), cf. 374c5–7), but in the cases of bodily conditions, if one has them willingly, one doesn't really have them, but rather the ability to act as if one had them (as Aristotle notes, *Metaphysics* 1025a12–13). Socrates' discussion of these cases invites the sort of correction Socrates made of Hippias' distinction between the liar and the truth-teller—both accomplish what they do by the same power. It's surely by the power of healthy voice, legs, and eyes, along with the power to fake a bad bodily constitution, that one "is willingly tuneless, lame, and poor-sighted". (Perhaps the point of this stretch of discussion is to raise the question of the difference between the kind of power a bodily condition is and knowledge or expertise.)

In Socrates' third group of cases—instruments such as eyes, a rudder, and animals we use, such as a horse—'willing' attaches to the human agent who uses them, but at the same time Socrates introduces the idea that the instrument in the case of an animal is its temperament or soul (*psuchê*) (374e–375a). By saying, 'better to have a horse whose soul allows one to ride it well or badly' (375a), instead of 'better to have a horse which can be ridden well or badly', Socrates paves the way for the idea that a human soul too is an instrument—better for the archer to have a soul with which he can miss the target willingly (375b). But 'willingly' and 'unwillingly' also easily attach to the soul as agent: the better soul is the one that errs willingly, in medicine, or music, and so on (375b–c). And when Socrates reintroduces 'we who possess a soul' to ask the question whether we'd rather have a soul that errs and does bad things willingly (375d) it is not clear whether to think of this on the model of having eyes that see poorly willingly (in which case they see well but we act as if they see poorly, or perhaps we temporarily obstruct their ability to see well). But who are 'we' over and above our soul? Or is the conclusion that we should prefer that we ourselves be able to err and do bad things when we want? But this raises the question: for the sake of what would I want to err and do wrong, not in running a race or shooting an arrow, but quite generally?

Whereas the *Gorgias* makes 'wanting to  $\phi$ ' and 'willingly  $\phi$ -ing' conditional on whether  $\phi$ -ing achieves the end for the sake of which one  $\phi$ -s, the *Hippias Minor* makes ' $\phi$ -ing wanting to/willingly' a matter of  $\phi$ -ing having been able to not- $\phi$  (374b), and contrasts ' $\phi$ -ing wanting to/willingly' with ' $\phi$ -ing from sickness (366c), luck, ignorance, for example.' This difference can explain why the *Gorgias* and *Hippias Minor* come to such divergent conclusions about the impossibility vs. superiority of willing wrongdoing despite their superficially similar conceptions of power as  $\phi$ -ing when one wants to. The *Gorgias*, as we saw in Section 3.4, argues for 'no one does wrong willingly, that is, wanting to' from (a) the assumption that people do wrong in order to be happy, (b) the claim that we do what we want, in the case of actions we do for the sake of some end, depending on whether our action actually is a means to our end, and (c) the argument that wrongdoing in fact is worse for the agent than even suffering wrong, and so detracts from the agent's happiness. By contrast, the *Hippias Minor* is silent on the question of why, for the sake of what, one might want to  $\phi$ . The expert  $\phi$ -er may have

the power to  $\phi$  at any time, but what informs whether, at a given time, he should  $\phi$ ? The *Gorgias* is able to answer this for each of the expertises by the hierarchical structure based on which expertise uses and judges the products of which (511c–512b, cf. 465c–d), with the health of soul at the top of this structure; by contrast, the *Hippias Minor* treats all expertises and all activities on all fours.<sup>65</sup>

This suggests that even though the desire for the good that is common to everyone doesn't define virtue in a way that distinguishes the virtuous from the vicious, a conception of virtue that simply ignores what is common to all will go astray. So I agree with the scholars who say that the conclusion of the *Hippias Minor* is contrary to fact, but I have argued that the fact to which it is contrary is the widely agreed 'everyone wants the good', rather than the 'Socratic paradox' no one does wrong willingly.

### 3.5 The Unwillingness of Badness from the *Republic* to the *Laws*

Section 3.4 argued that the *Gorgias*' account of instrumental desire underwrites Socrates' claim that actions undertaken for the sake of an end that undermine that end are unwilling because they are contrary to the agent's want, which is for the end; as a result, wrongdoing which is undertaken for the sake of happiness but in fact produces unhappiness is not wanted, but unwilling. Although a plausible explanation for such instrumental mistakes is ignorance—lack of intelligence (466d–e), lack of knowledge or power (509d–510a)—about what conduces to the end, it is not ignorance but contrariety to one's desire for the good that constitutes the actions' unwillingness. When we turn to the middle and late dialogues, we first see Plato extending the account of unwillingness as contrariety to the agent's desire for good from actions to conditions, including the condition of ignorance; and then attempting to distinguish voluntary from involuntary wrongdoing for the purposes of legislation in a manner consistent with the involuntariness of all wrongdoing.

In the *Republic* Socrates describes the condition of ignorance—that is, having false beliefs—as itself unwilling. He describes acquiring a false belief as (the result of) speaking falsely (*pseudesthai*) to the best part of oneself, and he says no one will willingly (*oudeis hekôn ethelei*) do this, at least about the most important things (382a). He also describes losing a belief as

‘expelling’ a belief (412e), which, he says, one may do voluntarily (*hekousiôs*)—when one learns that the belief is false—or involuntarily (*akousiôs*)—when the belief is true. These are voluntary and involuntary because people are voluntarily deprived of bad things, involuntarily deprived of good things, and beliefs are bad when false, good when true (413a). Socrates then lists the mechanisms by which one might undergo involuntary belief-loss: one might be persuaded by argument (*metapeisthentas*) or forget (*epilanthamenous*), which are cases of theft; one might be forced (*biasthentas*), which is when grief or pain causes a change of mind; one might be bewitched (*goêteuthentas*) by pleasure or fear (413b–d).<sup>66</sup> However, it isn’t the irrationality of these mechanisms that makes the loss of true beliefs involuntary; instead, Socrates emphasizes that it is a good thing we are losing, and we are involuntarily deprived of good things, treating the mechanisms simply as ways in which an involuntary belief-loss can come about.

The *Republic* doesn’t argue for the view that unwillingness is due to contrariety to the agent’s wanting, which is directed at his own good, presumably because the challenge Glaucon and Adeimantus have put to Socrates assumes the view (360d; Section 3.3.2 in this chapter). This is presumably also what allows Socrates’ ‘willing’ and ‘unwilling’ in the *Republic* to extend beyond instrumental actions to all actions and conditions.

But the Myth of Er that completes the argument of the *Republic* raises a problem for the view that contrariety to the agent’s want for his own good is sufficient for unwillingness. In the afterlife souls choose their next lives, and some of their choices are of bad lives. Socrates emphasizes that the choosing souls are responsible (*aitios helomenou*, 617e)—for God cannot be responsible for bad things (cf. 380b–c). How can a soul choose, or be responsible for, something it does unwillingly?

The *Timaeus* recognizes the problem and suggests that some kind of forward-looking responsibility for improvement may be compatible with our past bad actions having been unwilling. Timaeus says that no one is bad willingly (*kakos . . . hekôn oudeis*, 86d7–e1), and explains that badness is due to two causes: first, bad bodily conditions—excessive pleasures and pains that make the soul unable to see or hear aright or share in reasoning (86c2–3) and make the reasoning part of the soul forgetful and hard of learning (87a7)—and second, bad constitutions with bad culture and education (87b1–3). These he calls ‘most involuntary’ (*akousiôtata*, 87b4<sup>67</sup>), adding that our begetters and nurturers rather than we ourselves

are responsible (*aitiateon*) for our badness—although we should try to improve our characters in whatever way we can.<sup>68</sup> But since our begetters and nurturers are responsible whether we turn out good or bad, the feature that makes vices but not virtues unwilling conditions seems to be, again, that the vices are ‘hateful/hostile conditions’ that ‘come to be for everyone unwillingly’ (*panti de tauta echthra kai akonti prosignetai*, 86e2–3)—they are unwilling precisely because they are bad for us. Now in context, Timaeus tries to show that vice is folly (*anoia*)—either madness (*mania*), or ignorance (*amathia*, 86b3–4), which includes weakness (*akrateia*) before pleasures (86d–e)—the responsibility for which lies with the body and society.<sup>69</sup> Such a ‘disease of the soul’ is in need of treatment, which consists in various kinds of motion to restore the correct proportion between soul and body (87c–89d), rather than blame.

Finally, in the *Laws*, Plato confronts the fact that deeming all wrongful actions and bad states of character involuntary provides no guidance to the law, for which the distinction between voluntary and involuntary is fundamental. The Athenian begins the discussion by reiterating that no one is bad or unjust willingly (860d, cf. 731c: the unjust person is unwillingly unjust, for no one ever acquires bad things willingly, least of all into his soul, which is the most valuable part of himself. For this reason, we should pity the unjust person, calming our anger against him, if he is curable). I’ll first translate the Athenian’s speech and enumerate the key points before commenting on the argument.

[Athenian:] . . . (L1) [T]he bad are in all respects unwillingly bad, and this being so, the argument next in order follows necessarily: . . . (L2) the unjust person is bad I suppose, and the bad person is of such a sort as to be unwilling I suppose. (L3) But it isn’t reasonable for [a] voluntary [thing] to be done involuntarily, (L4) so [*oun*] to the person who posits an involuntary injustice (*tên adikian akousion*), and indeed, this ought now to be agreed to by me, the unjust person would seem to do injustice unwillingly. For I agree that all do injustice unwillingly—and if someone says that while some (*men*) are unwillingly unjust for the sake of ambition or love of honour, still (*mên*), many do injustice willingly, which that account of mine [says], but not this—in what way could I harmonize my [own] accounts? . . .

Will you distinguish . . . involuntary and voluntary acts of injustice (*adikêmata*), and shall we lay down greater punishments for the voluntary errors and acts of injustice and lesser for the others? Or [shall we lay down punishments] equally for all of them, as not being voluntary acts of injustice at all? . . . (860d1–861a2)

. . . [I]t is necessary before legislating somehow to clarify that there are two [kinds, viz. voluntary and involuntary] and the difference [between them] is of

another sort, in order that, whenever someone lays down the penalty for either [a voluntary or involuntary act], everyone follows the arguments, and is able to judge whether the judgement is in some way laid down fittingly or not . . . (L5) [I]t seems [that] in communities and associations many harms [committed] by citizens against one another come about, and in these [cases] it is unobjectionable [to say] 'voluntary' and 'involuntary' . . . But no one should posit that all these harms are injustices, and thus think that all injustices are in themselves two-fold, some voluntary, and others involuntary—for of all of them the involuntary harms are lesser neither in number nor in magnitude than the voluntary . . . (L6) For I will not say . . . that if someone does some damage to someone not wanting/intending to but unwillingly (*mê boulomenos all' akôn*), he does injustice, but unwillingly, and I will not legislate in this way, legislating as if this is an involuntary act of injustice, but I will posit that this sort of harm, whether it is a greater or smaller [harm] for the one who suffers it (*tôi gignêtai*), isn't injustice at all. And if my position prevails, [then] we will say that often, when an incorrect benefit comes about, the person responsible for the benefit does injustice. For, my friends, neither if someone gives something to someone else, nor, on the contrary, if he takes it away, should one say that such a thing is thus simply just or unjust, but (L7) if he harms or benefits someone else with (*chrômenos*) a just disposition and manner, this is what should be seen by the legislator, and (L8) he should look at these two [things], both the injustice and the harm, and first (*men*) the harm should be made healthy as far as possible by the laws, and the thing being destroyed should be preserved . . . [.] and second (*de*) the atonement by compensations having been made, the laws should always attempt to establish friendship from difference between the doers and the sufferers for each of the harms. (L9) As for unjust harms and gain, if someone doing injustice makes some gain, as many of these as are curable should be cured, since they are diseases in the soul. But one should speak of the direction (*to rhepein*) of the cure of injustice in the following way . . . In order that the one who acts unjustly (*adikêsêi*), in a great or small respect, the law will teach him and will compel him entirely to willingly dare to do such a thing either never again or much less often, and in addition to this, he will pay the penalty for the harm. (861c2–862d4)

In (L1) and (L2), the Athenian maintains that badness (the condition, as in *Timaeus*) is unwilling and so injustice, being a form of badness, is unwilling. In (L3), he derives the consequence that all unjust actions are involuntary, for 'something voluntary can't be done involuntarily'. By 'involuntarily' here I understand 'from something involuntary in the agent'—taking as my model the *Protagoras*' explication of acting foolishly in terms of acting from folly and acting temperately in terms of acting from temperance (332b, cf. Chapter 2.2.3). So (L3) means: if an act is to count as voluntary, its cause or origin can't be something involuntary in the agent—which the condition of injustice is.

(L4) now draws out a consequence that applies to those who, like the thinkers discussed in Sections 3.3 and 3.4, consider some or all injustice involuntary: these thinkers must say (as must the Athenian) that the doer of injustice does injustice unwillingly. How then is the Athenian to legislate, since legislation depends on distinguishing between voluntary and involuntary acts and treating the voluntary agent differently from the involuntary? (L5) proposes applying the voluntary/involuntary distinction not to justice and injustice, but to acts described as benefits or harms. Thus it becomes possible to say that (L6) a person who harms someone not wanting/intending to, but unwillingly, isn't doing an injustice at all, and in such cases, (L8) what the law must do is to recompense the one who has been harmed. (L7) explains that an act counts as just or unjust in virtue of the just or unjust character of its agent, which is the cause or origin of the act. The Athenian will go on to describe the condition of an unjust character as one in which anger, fear, pleasure, or appetite overcomes the judgement of what is good (863e–864a). Given its concern to improve the character of the unjust, in the case where the harm is also an act of injustice, the law must (L9) treat the agent of injustice in whatever way necessary (teaching, compelling, etc.) to cure him of his injustice.

Unfortunately, the Athenian leaves two crucial points unspecified. First, since an intentional harm may be just (e.g. if it is a punishment for some crime) and an intentional benefit may be unjust (e.g. if it is a bribe), by what criterion are acts of harm and benefit determined to be just or unjust? (L7) says by the disposition of the soul, but dispositions have to be inferred from actions. The Athenian must be supposing that the criterion is legality, so that illegal actions are taken to be evidence of disregard for the law, and the disposition to disregard the law is injustice.

Second, the Athenian does not say what criteria are to be used for determining whether these (illegal) acts of harm and benefit are done voluntarily or involuntarily. In 'Injury, Injustice, and the Involuntary in the *Laws*', Malcolm Schofield argues that on the *Laws*' account, acts of unjust injury must be 'in a way voluntary, in a way involuntary[:] . . . involuntary because they are the sorts of act nobody would choose to do if they understood the extreme harm the injustice that is their source wreaks upon their perpetrator's soul[:] . . . voluntary because they are injuries



deliberately committed against others in full knowledge that they are wrongful'.<sup>70</sup> In other words, the same act may be voluntary qua injury but involuntary qua injustice.<sup>71</sup>

The problem with Schofield's account is that what makes a harm voluntary or involuntary seems to be pretty much the same as what makes injustice voluntary or involuntary, namely ignorance. So qua act of injustice, the act is involuntary because the agent doesn't know the harmfulness of his act to himself, and qua injury, the act is voluntary because the agent does know the wrongfulness of his act. Yet the Athenian says the distinction between voluntary and involuntary he is introducing for legislation is different (*tên diaphoran allên*, 861c3) from the distinction between voluntary and involuntary according to which all injustice is involuntary. By contrast, on my account of 'unwilling' as 'contrary to natural desire for the good', the two distinctions are indeed very different. Injustice is unwilling because it is contrary to our natural desire for what's good for us; however, for the purposes of legislation, we can speak of a voluntary act of injustice just in case an agent knowingly commits an act of harm or benefit contrary to the law (and we do not recognize any such thing as an involuntary act of injustice).<sup>72</sup> This second distinction between the voluntary and the involuntary, incidentally, is what Socrates used in the *Apology* passage we examined in Section 3.2, when he argued that if he corrupted his associates unwillingly, that is unintentionally, Meletus should have instructed and exhorted him rather than prosecuting him in court.

But Schofield is right to emphasize that on the Athenian's view, the same act is voluntary qua injury and involuntary qua injustice. This is a problem, for it gives us two conflicting criteria for the voluntariness or involuntariness of a single action. Take an act of injustice/intentional illegal injury: in reality, since something voluntary doesn't come to be from something involuntary (L3), and people are bad and unjust unwillingly ((L1) and (L2)), this action is involuntary; in law, however, that same action must be voluntary if it is unjust, and its voluntariness is established by the agent's intention. Perhaps the Athenian has said enough to justify different treatment under human law (treatment or punishment for those whose character is deemed to be the cause of their illegal injury, on the basis of their knowledge of its illegality), but is it enough to justify his characterizing the act, flowing as it does from an involuntary state, as 'voluntary'? And what of divine justice? Does it

ignore the distinction between voluntary and involuntary, and simply treat involuntarily unjust souls directly rather than on the basis of their actions? Can Plato stably hold that divine justice involves punishing the involuntarily bad on the grounds that punishment is simply cure?

## Appendix: Voluntary Action and Character in *Nicomachean Ethics* III.5

Plato's discussions of the voluntariness of our actions and character-states illuminate Aristotle's famous and obscure arguments about the voluntariness of our character-states in *Nicomachean Ethics* III.5. While it's widely understood that Aristotle's arguments are responding to Plato's view that our vicious actions and character-states are involuntary, but not our virtuous actions and character-states, and that Aristotle's basic response is to argue for symmetry in what we say about virtue and vice, and virtuous and vicious actions, we are now in a position to be more precise. Aristotle builds up to his conclusion in *NE* III.5 that not only are our virtuous acts and states of character voluntary, but also our vicious actions and states of character, by stating necessary conditions for actions to be involuntary in *NE* III.1: they must be compelled, or done due to ignorance of the circumstances of action and accompanied by pain (anticipatory or retrospective). While these negative characterizations may be stated positively—voluntariness requires the action to have its origin in the agent, and requires the agent to know what she is doing—they are an indication that Aristotle has in mind the tradition of argument that has focused on the involuntary. In III.2–4, Aristotle defines choice, the voluntary attitude we have towards the actions we undertake as means to our ends, as arising out of deliberation about how to bring about the end we want, the end being an apparent good. Aristotle begins III.5 with a summary of these accounts: given what he's said about what is wanted, and what is deliberated about and chosen, it follows that actions from (*kata*) choice are also voluntary. And the exercises of the virtues are also concerned with these (actions from choice) (1113b2–6). None of this should be controversial, although Aristotle has (as noted in Section 3.4) restricted the term 'wanting' to the apparent good that is the agent's end (as represented) and assigned the term 'choice' to our attitude towards actions done as means to our ends.

What *is* controversial is the conclusion of Aristotle's first argument, 'virtue is up to us, and also vice' (1113b6–7). Aristotle argues:

- (NE1) Where  $\phi$ -ing is up to us (*eph' hêmin*), not- $\phi$ -ing is up to us also, and vice versa (1113b7–8).

- (NE2) So where  $\phi$ -ing, when  $\phi$ -ing is something fine, is up to us, so is not- $\phi$ -ing, when not- $\phi$ -ing is something shameful, and vice versa (1113b8–11).
- (NE3) And if doing fine and shameful things, and likewise not doing them, is up to us, and if this is being good or bad, then being decent or base is up to us (1113b11–14).

As previous commentators have noticed, Aristotle doesn't accept the equivalence between doing fine or shameful things and being good or bad (NE3). Susan Sauvé Meyer explains (NE3) as due to Aristotle's dialectical opponents being proponents of the Socratic paradox, who sometimes express 'no one does wrong willingly' 'using terminology that might equally well refer to states of character'.<sup>73</sup> Susanne Bobzien argues that since Aristotle hasn't yet shown that virtue and vice are up to us qua character states, by 'virtue' and 'vice' he must mean characteristics manifested in actions to talk about agents qua agents of actions.<sup>74</sup> But as we have seen in Section 3.5, 'no one does wrong willingly' and 'no one is bad willingly' are not different expressions of the same thesis, but distinct theses that may be connected in different ways. And it seems to me unlikely that Aristotle, after explaining how actions are the causes of our character states in *NE* II, and right before using this causal thesis to argue that our character-states are voluntary in III.5, can just *mean* by 'virtue and vice', 'virtue and vice as manifested in actions'.

On the other hand, if Aristotle's dialectical opponent is not Socrates or Socratics in general, but the Athenian of *Laws* IX in particular, we have an explanation for not only the un-Aristotelian equivalence between doing fine and shameful things and being good or bad in (NE3), but also the (anti-Athenian) claim in (NE2), that if it's up to us to act when it's fine to act, it's up to us not to act when it's shameful not to act. Both Aristotle and Plato's Athenian accept (NE1) for actions, insofar as they are actions that benefit or harm. (NE2) moves from actions thus characterized neutrally to actions characterized evaluatively, contrary to the Athenian's judgement that the wrongful actions would be involuntary—but using the non-theory-laden expression 'up to us'. (A2) forces the issue Schofield identified in *Laws* IX, that the same action is, qua injury or benefit, voluntary, but qua injustice, involuntary. Aristotle is arguing that if the criterion of 'up to us' for  $\phi$ -ing is 'we can  $\phi$  or refrain from  $\phi$ -ing' as in (NE1), and if  $\phi$ -ing is up to us when  $\phi$ -ing is something fine, how could it fail to be up to us to refrain from  $\phi$ -ing only when refraining from  $\phi$ -ing is base? How in particular could *that* description change an action from voluntary to involuntary?

As noted above, Aristotle rejects the equivalence in (NE3). Aristotle's difference with Plato lies in his extension of the (legal) voluntariness of neutrally characterized actions to evaluatively characterized actions. What enables him to do this is that in III.1 he has given perfectly general criteria for determining when an action is voluntary (the principle of the action is (in) the agent and the agent

knows the particulars) that do not make any reference to the agent's character, and in III.2–4 he has explained that character is expressed in (and so is better revealed by) chosen, rather than merely voluntary, actions. The distinction between choice and the voluntary allows him to break with Plato's supposition that an action that is unjust must be so because it flows from an unjust character—which, since Plato supposes the unjust character to be involuntary, would make the action involuntary. No, Aristotle is saying, there are unjust actions that are chosen that come from an unjust state of the soul, but there are also unjust actions that are merely voluntary—neither compelled nor done because of ignorance of the circumstances—and indeed, the soul comes to be in an unjust state by doing such unjust actions. In II.4, anticipating the (Platonist) objection to his account of habituation that we cannot become just by doing just actions because we must already be just in order to do just actions (1105a17–21), Aristotle explains that actions are called just when they are of the sort the just person would do, but a person is just when he not only does these things, but does them in the way a just person does them (1105b6–10)—in other words, knowingly, choosing to do them for their own sake, and stably (1105a28–33).

But this is not Aristotle's most precise statement of the relationship between just act and just character. In *NE* V.7–8, he distinguishes 'what is unjust' (*to adikón*), which he defines as unjust by nature or ordinance, and 'an act of injustice' (*adikêma*), which he defines as what is unjust when it has been done (1135a8–13). He then explains that a person 'does injustice' (*adikei*) when he voluntarily (*hêkon*) does what is unjust. (Aristotle doesn't seem to distinguish between *adikêma* and *adikein* except insofar as the former refers to the thing done and the latter to the doing of it, both of which require voluntariness.) If he does injustice involuntarily, he doesn't do injustice (*adikei*), or does injustice only coincidentally (*kata sumbebêkos*). Similarly, an act of injustice is defined by the voluntary and involuntary (*tôi hekousiôi kai akousiôi*) (1135a15–20). When something unjust is done voluntarily, it is blamed, and at the same time is also an unjust act; however, there may be something unjust but no act of injustice, when the voluntary is not present in addition (1135a20–23).

Whereas *Laws* IX defined 'unjust act' as 'violation of the law flowing from an unjust state of soul', Aristotle defines it as 'voluntary violation of the law'. This definition allows Aristotle to preserve Plato's judgement that there is no 'involuntary injustice', not simply by denying the injustice of anything involuntary, but also by distinguishing between the conditions for voluntary and chosen (viz. flowing-from-character) action. For as long as what one does from among the things that are up to one, one does knowing all the particulars of the action so that one does it non-coincidentally,<sup>75</sup> and one is not forced, one's action is voluntary (1135a24–b2). These conditions are occurrent attitudes of the agent rather than stable states of character. (Of course a stable state of character would

give rise to certain sorts of occurrent attitudes, but we need not suppose that whatever gives rise to occurrent attitudes at any moment is itself stable.) Even when a person inflicts an unjust injury knowingly, if he does so without previous deliberation (*mê probouleusas*), but, for example, out of anger, he does injustice and it is an act of injustice that he does, but he is not thereby an unjust person. It is only when a person inflicts an unjust injury as a result of deliberation, forethought (*ek pronoias*) and choice (*ek prohaireseôs*) that he is an unjust person (1135b19–1136a3). The relationship between something unjust (*to adikón*) and an unjust agent (*ho adikos*) is far more complicated than Plato thought, for an unjust person is one who does injustice (*adikei*) and commits an act of injustice (*adikêma*) not only voluntarily, but also from deliberation and choice.

Aristotle's second argument in *NE* III.5 addresses the core of Plato's position that voluntariness is asymmetrical as far as virtue and vice are concerned. On Aristotle's view, 'no one is unwillingly happy, but vice is voluntary' (1113b14–17). Presumably no one is unwillingly happy because whether, like Plato, you posit an attitude of wanting which everyone has towards being happy, or, like Aristotle, you regard being happy as the *telos* of a human being which one may represent correctly or incorrectly in wanting (*boulêsis*), being happy will be in conformity with at least *an* internal impulse of any agent. Before turning to Plato's argument for the involuntariness of vice, Aristotle raises considerations in favour of his own view that vice is voluntary: otherwise we won't be the origins of our actions (1113b17–21). But the law treats us as the origins of our actions, disciplining and punishing us for our wicked actions as long as we aren't (causally) responsible for them through force or because of ignorance (about the particulars of our action) (1113b23–25), punishing even in the latter case if the agent is (causally) responsible for the ignorance because of which he does wrong (e.g. by getting drunk and then injuring someone because he doesn't know who he is, or what he's doing) (1113b30–33). Aristotle overrides the objection that the origin of action is the agent's character, which may be, for example, such as to not take care (*viz.* to ensure that he knows what he should, e.g. what the law is), by pointing out that we acquire our characters by the actions we perform, a fact which everyone knows: that's why athletes and musicians practise for their competitions and performances<sup>76</sup> (1114a3–10).

Now of course there's a difference between the athletes and musicians who perform the actions they do in order to acquire athletic and musical abilities, and the unjust person, who isn't performing unjust acts *in order to* become unjust. And this takes us to the heart of Plato's reason for maintaining that vice is involuntary: it is bad for us, and contrary to our desire for our good—from which it follows that vicious actions, since they either lead to (*Gorgias*) or arise out of (*Laws*) an involuntary state, must also be involuntary. So Aristotle's next claim is that it's unreasonable to say, as Plato would, that the doer of injustice doesn't

want (*mê boulesthai*) to be unjust (1114a11–12). Voluntariness doesn't require not conflicting with *boulêsis* or any other single principle in the agent (that's how akratic action, contrary to the agent's want, is still voluntary); for voluntariness, it suffices that the agent is not compelled (for compulsion, there would have to be *no* impulse towards the action from the agent) and knows the particular circumstances of his action. So, for example, suppose I know that by stealing this lipstick in the convenience store, and cheating on that math test, and so on, I'm going to end up becoming a dishonest person. I want to steal and cheat insofar as these get me a lipstick and a 'B', but I don't want to steal and cheat insofar as these get me a dishonest character. Aristotle's point is that it doesn't matter that I don't (as I'm stealing and cheating) want to become a dishonest person, or that once I am dishonest, I will wish I weren't, or that I will not at that point be able to be honest even if I want to be: my dishonesty (my state) is still voluntary, for it developed out of my voluntary actions and I knew it would (1114a12–21).

Socrates in the *Republic* and Timaeus in the *Timaeus* had proposed accounts of the external causes by which we might come to have and do bad things (Section 3.5). For Timaeus this was part of the build-up to likening vice to bodily illness, in need of treatment. Aristotle now argues that insofar as we are, through our voluntary actions, the causes of our bodily illnesses or defects (e.g. by not exercising we end up being weak in body), our illnesses and defects too are voluntary—evidence of which is that these are blamed (1114a15–31).

Aristotle's last arguments in *NE* III.5 respond to a hypothetical (Platonist) objector who argues on Aristotle's own grounds.<sup>77</sup> First (the objector argues), if we all pursue the apparent good, and what appears good to each depends on the sort of person he is (as Aristotle argued in *NE* III.4), then we are not masters of our appearances (1114a31–33). Aristotle replies that either (1) we are the causes of the sorts of persons we are and so are in a way causes and thus masters of our appearances, or (2) we are not, and no one is the cause of his doing bad things, but instead, knowledge (or ignorance) of, and the (in)ability to correctly aim at, the end will be something we are born with, like eyesight, and no one will be able to learn it from another, or to develop it in any other way. But in this case, virtue will be no more voluntary than vice, and this goes for actions as well as states (1114b1–16). But this is as unacceptable to the Platonist as to Aristotle. So (1) either how the end appears to each of us is to some extent dependent on us, or else, (2\*) while the end is natural, virtue is voluntary by the good person 'doing the rest' (presumably, actions for the sake of the end) voluntarily. But then (on (2\*)) vice will be no less voluntary than virtue, for 'through himself' (*di' hauton*) may be said of the bad person's actions, even if not of his ends (i.e. as may be said of the good person's actions) (1114b16–21). And if the virtues are voluntary because we are co-causes of our characters, then so are the vices, for the two come about in similar ways (1114b21–25).

## Notes

1. This reconstruction is based on G. Santas' influential 'The Socratic Paradoxes', *Philosophical Review*, 73 (1964) [= 'Socratic Paradoxes'], 147–64.
2. Why so many terms? J. L. Austin's 'A Plea for Excuses' (*Proceedings of the Aristotelian Society*, ns. vol. 57 (1956), 1–30) makes the general point that adverbs and their negations do not apply to verbs symmetrically, because many verbs are said *haplôs*. This would explain why in English 'voluntarily' and 'involuntarily' are not strict opposites, but the opposite of 'involuntarily' is 'deliberately' or 'on purpose', and the opposite of 'voluntarily', 'under constraint' (17), and why there is such a thing as doing something 'inadvertently' but not 'advertently'. Presumably it also explains why the opposite of 'intentionally  $\phi$ -ing' is not 'unintentionally  $\phi$ -ing', where that would mean ' $\phi$ -ing by accident'.
3. According to the so-called 'Knobe effect', the conditions under which an action A is said to be done 'intentionally' are different depending on whether the actual consequences are good or bad. If the actual consequences are bad, A is said to be done 'intentionally' even if the agent only had foreknowledge of the bad consequences and intended to bring about some other consequence. If the actual consequences are good, then in addition to foreknowledge, the agent needs also to have intended to produce them (J. Knobe, 'Intentional Action in Folk Psychology: An Experimental Investigation', *Philosophical Psychology*, 16 (2003), 309–24). If this is right, then what Socrates says here revises our ordinary judgements of when an act is done intentionally in one direction, since he requires 'intending to do it' for both bad as well as good consequences, and when Aristotle argues against his position (discussed in the appendix to this chapter), he revises ordinary judgements in the opposite direction, arguing that we need not intend/will something to do it intentionally/willingly, but foreknowledge is sufficient.
4. Socrates' apparent endorsement of conditions in which the law requires wrongdoers to be brought to court and punished (26a) suggests that if he could be shown to corrupt his associates willingly, punishment would be appropriate.
5. This puts us in a position to reconsider the possible psychological eudaemonism of Gorgias' *Defence of Palamedes*, noted in Chapter 2. According to Calogero, 'Gorgias and the Socratic Principle', in the *Defence*, Gorgias anticipates the Socratic principles 'no one does anything unless he prefers it to other things he can do' and 'no one does wrong willingly'; Calogero also argues that Socrates' defence is modelled on 'Palamedes'. I have argued above that 'no one does anything unless he prefers it to other things he can do' is not, generally speaking, a Socratic principle. Is it a Gorgianic one? In the *Defence*,

'Palamedes' says that people do everything they do for the sake of one of two things, to obtain a benefit or to flee a harm—unless they are mad (19). 'Palamedes' first argues that he couldn't have betrayed the Greeks as accused, because circumstances would have made it impossible for him to perform the act (6–12). He then enumerates possible ends for the sake of which he might have performed such an act: being a tyrant (13), getting wealth (15) or honour (16), helping friends and harming enemies (18), and he argues against each of these: wealth is not something he would want given his moderate expenses and adequate means; likewise, he already enjoys enough honour; treason could not secure tyranny and he lacks the power to be tyrant; finally, treason actually results in dishonour, and helps enemies and harms friends. This line of argument culminates in his saying that his accuser inconsistently charges him with cleverness (*sophia*) and madness (*mania*) (25). 'Palamedes'' strategy is to argue that if he's clever (clever enough to pull off the treason), then he can't also be performing an action so antithetical to any of the ends he might reasonably have that if he were, he'd be mad. The reasoning bears on what intentions can be reasonably attributed to him.

6. Santas, 'Socratic Paradoxes', argues that we should not 'lump together' the claims, 'no one does bad willingly' and 'no one does wrong willingly' on the grounds that Socrates assumes and never argues that the bad is such as to harm, but must argue, e.g. against Callicles, that acting wrongly/unjustly harms while acting justly benefits the agent (149–50). But at least here it is Socrates who lumps them together.
7. Simonides is not alone in saying that wrongdoing can be involuntary. Cf. Euripides *Augè* 265.1: *to d' adikêma egenet' ouch hekousion*. And in *Republic* I Cephalus says that that wealth is valuable to a decent person because the possession of wealth plays a big part in his not cheating or deceiving anyone against his will (*akonta*) (331b). Thanks to Clerk Shaw for drawing my attention to this passage.
8. This interpretation follows A. Beresford, 'Nobody's Perfect: A New Text and Interpretation of Simonides PMG 542', *Classical Philology*, 103 (2008), 237–56. But I don't accept Beresford's relocation of the lines beginning with 'only god can have this prize' to the first stanza. This is because Socrates takes 'this prize' to refer to Pittacus' words '*esthlon emmenai*', and makes a great fuss about the difference between Pittacus' '*emmenai*' and Simonides' '*genesthai*' in the first stanza, with the result that 'only a god can have this prize' must come *after* 'being good is hard'. Hayden Pelliccia suggests that 'this prize' that belongs to the god is its being difficult (as opposed to impossible) to be good. (I would like to thank him, Tad Brennan, and Charles Brittain for discussions about Simonides' poem and Socrates' interpretation.)
9. The manuscripts disagree between *euêtheias* and *eunoias*.



10. To explain the translation of the substantive as 'liar' but of the verb of action as 'speak falsehood': Hippias faults Odysseus for being a willing speaker of falsehood—what we would call a liar—and Socrates then zooms in on the action of speaking falsehood and questions whether it is better to do this willingly or unwillingly, because if the answer is 'willingly', then Hippias has lost his stated reason to prefer Achilles to Odysseus.
11. Gorgias is concerned with compulsion and force, and does not use the terms 'willing' and 'unwilling' (*hekôn/akôn*). An anonymous reader has pointed out that in section 15 he uses the term 'wish' (*thelomen*): 'that which we see has a nature, not the one we wish, but whatever it turns out to be'. Still, Gorgias uses the contrast between what we wish to see and what we actually see to bring out the passivity of our response to the objects of perception (and hence to perceived beauty). For the case he is making throughout is that Helen was compelled by a power stronger than herself.
12. Supplied by Diels (see T. Buchheim (ed. and comm.), *Reden, Fragmente und Testimonien/Gorgias von Leontini* (Hamburg: F. Meiner, 1989)).
13. M. Gagarin and P. Woodruff (trans.), *Early Greek Political Thought from Homer to the Sophists* (Cambridge, 1995) [= *Early Greek Political Thought*], 191–4.
14. Gagarin and Woodruff, *Early Greek Political Thought*, 245.
15. In *Socratic Paradox*, Roslyn Weiss argues that 'the noxious notion that no one is willingly just' (12) is maintained not only by those mentioned in Glaucon and Adeimantus' challenge, but also in one or another version by Protagoras, Polus, Meno, Hippias, etc. (9). But there is a greater diversity of opinion among these intellectuals than this, and Socrates recognizes it.

Weiss also argues that Socrates cannot literally believe that no one does wrong willingly, for (1) if he were not aware that some people 'deliberately choose to commit injustice... it would be difficult to understand why he considers some people—and not others—wicked' (23), and (2) Socrates himself is a man who does wrong only unwillingly (142, 147). But against (1), Socrates can distinguish the good from the bad in terms of what they do, rather than whether they do it willingly or unwillingly, and against (2), Socrates does not consider himself good by this standard. Socrates does invoke the standard 'don't do wrong willingly' occasionally, e.g. when he asks Crito, 'Do we say that one must never in any way do wrong willingly (*hekontas adikêteon*)?' (*Crito* 49a). But this is the beginning of an argument to convince Crito that Socrates should not escape prison as this would be unjust, viz. destructive of the city so far as possible, so perhaps Socrates is starting from Crito's standards. Or perhaps this is another case (like *Apology* 25c–26a) where *hekôn* is best

understood ‘intentionally’, and the starting point is that one should not do wrong if one knows it’s wrong. Another example is his saying at his defence, ‘I do not willingly wrong anyone (*hekôn . . . adikein*)’ (*Apology* 37a–b). But Socrates’ not willingly wronging anyone is consistent with no one’s willingly wronging anyone.

Weiss is not wrong to understand Socrates as responding to the amoralist challenge, but a too-narrow focus on the amoralist challenge leads us to overlook what else Socrates is responding to, what he shares with the amoralists and why, and as a result we miss what he is contributing to his own intellectual world.

16. On Euripides’ philosophical outlook, Diogenes Laertius’ *Life of Socrates* (II.18) reports that the comedians Mnesilochus, Callicles, and Aristophanes lampooned Euripides for being helped by Socrates to write his plays. D. Sansone, ‘Plato and Euripides’, *Illinois Classical Studies*, 21 (1996), 35–67, documents Plato’s extensive quotations from Euripides.
17. Gagarin and Woodruff, *Early Greek Political Thought*, 229–36.
18. Plutarch reports that when an athlete unintentionally (*akousiôs*) killed one Epitimus with a javelin, Protagoras and Pericles spent a whole day puzzling over whether, according to the most correct account (*orthotaton logon*), the javelin, javelin-thrower, or organizers of the contest, were more responsible (*mallon . . . aitious*). In *Pericles* 36.3, C. Lindskog, and K. Ziegler (eds), *Plutarch: Vitae Parallelae* (Leipzig, 1957).
19. G. Rickert, *HEKÔN and AKÔN in Early Greek Thought*, *American Classical Studies* 20 (Atlanta, 1989), argues that in literature from Homer to Euripides, the terms ‘willing’ and ‘unwilling’ are used to characterize the agent’s overall attitude to his or her action, as one of deep commitment or divorcement. This attitude is not a simple function of whether the agent acted as she did because of force (*bia*) or compulsion (*anangkê*), for an agent may do what is forced or compelled willingly, e.g. although Helen was taken by force, she was taken willingly (*hekousia*, Euripides *Electra*, 1065–8). Rickert’s is meant to be an account not of what the terms mean, but of the conditions under which they are applied. It’s useful to distinguish Rickert’s project, which is to try to understand these terms from their contexts of use, from what Gorgias and Plato are doing (and I am trying to describe), which is to provide an account of what objective conditions make an action unwilling, even if that entails revising the use of terms.
20. I suspect that it is Gorgias’ conception of power that explains his unobjecting assent to Socrates’ saying at 460b–c that just as a person who has learned carpentry is a carpenter and a person who has learned music is a musician, so too a person who has learned justice is just, and will necessarily want to do,

and do, just things. Many commentators have objected that this relies on assumptions of Socratic moral psychology (viz. psychological eudaemonism) that Gorgias has been given no reason to accept (e.g. T. Irwin, *Plato: Gorgias* (Oxford, 1979) [= *Gorgias*], 126–7; J. Cooper, ‘Socrates and Plato in Plato’s *Gorgias*’, in *Reason and Emotion* (Princeton, 1999), 29–75 at 43–6). But why does Socrates have to give Gorgias any grounds to accept this? Gorgias has said that the orator’s power to rob other experts of their reputations gives him no reason to do it and that the orator shouldn’t do it but should use his power justly (457a). If Socrates reasons, what kind of power enables him to use his power justly, if not justice, then there is nothing particularly ‘Socratic’ about his reasoning. These commentators seem to think Gorgias should immediately object, ‘Justice is not an expertise as are music and carpentry! Justice combines knowledge with a disposition of one’s desires!’ But should Gorgias be able to say all this? Aristotle, writing with the benefit of studying Plato, could do it, but getting the desire component of virtue right seems to me a considerable achievement, and (as I hope to show in this chapter) Plato takes all of the *Gorgias*, *Meno*, and *Hippias Minor* to do it.

21. Socrates’ ‘so to speak’ acknowledges that he is using the words ‘doing what one wants’ in some special way (although, as I’ll argue shortly, he is not simply defying ordinary language and stipulating that ‘wanting’ must be directed towards what is actually good).
22. ‘Doing what one wants’ is popularly the prerogative of tyrants; see Herodotus III.80.3, where Otanes argues against monarchy that it permits the ruler to *poieein ta boulêtai aneuthunôi*. Thanks to Stephen Menn for pointing this out to me. Polus is claiming that this power belongs equally to orators.
23. At 504d Socrates says a good and expert orator will always *noun echôn* how justice may come to be in citizens’ souls.
24. For this complaint, see, e.g. Irwin, *Gorgias*, 141–5. Cf. K. McTighe: ‘What if the medicine were sugar-coated?’ (‘Socrates on Desire for the Good and the Unwillingness of Wrongdoing: *Gorgias* 466a–468e’, *Phronesis*, 29 (1984) [= ‘Socrates on Desire’], 193–236 at 204).
25. Homer uses *boulomai* to express a preference or choice (*Iliad* 1.117, 23.594; *Odyssey* 11.489, 12.350). But probably it’s the hierarchy implied in doing/wanting to do one thing for the sake of another rather than the connotation of *boulesthai* that is doing the work here. A parallel would be Diotima: love is not of the beautiful, but of giving birth in beauty (*Symposium* 206e2–5). Aristotle makes the point clearly: desire or knowledge can be said to be ‘of’ the end, or of something towards (*pros*) that end (*Topics* 2.3, 111a1f.); the most correct thing to say about the object of a relation/thing said *pros ti* (6.3, 140b27) is that it is *of/pros* the end; e.g. desire is not for the pleasant but for pleasure, since we choose the pleasant for the sake of pleasure (6.8, 146b9–12).

26. Irwin takes Socrates to be assuming that when I want something it is because I believe it promotes my welfare, and when I choose one thing over another, it is because I believe that what I choose contributes more to my overall good (*Gorgias*, 142–3).
27. ‘Beneficial’ may be a word that marks the productive-of-goodness character that makes our attitude towards these instrumental, cf. ‘good’ at *Hippias Major* 296e.
28. Irwin regards (G1) and (G5) as conflicting and says that Socrates should reject (G1) in favour of (G5), for I may want x both as a means to y and for its own sake (*Gorgias*, 144–5).
29. Contra D. O’Brien, *The Socratic Paradoxes and the Greek Mind* (Chapel Hill, NC, 1967), 89, who claims, contrary to (G5), that Plato restricts *boulesthai* to ends. D. Wolfsdorf, although he notes that this argument is concerned with actions that are means to ends, succumbs to the same temptation when he rephrases ‘he doesn’t do what he wants’ as ‘he doesn’t do what he terminally desires’ and ‘he fails to satisfy his terminal desire’. See ‘*Gorgias* 466a4–468e5: Rhetoric’s Inadequate Means’, *Classical Philology*, 103 (2008) [= ‘Rhetoric’s Inadequate Means’], 103–34 at 128. Such silence about the attitude of the agent *to his action* loses the connection with the claim Socrates will want to make, that such an agent is acting unwillingly.
30. T. Penner, ‘Desire and Power in Socrates: The Argument of *Gorgias* 466a–468e’, *Apeiron*, 24 (1991) [= ‘Desire and Power in Socrates’], 147–202, argues for (G5\*) on the grounds that (Socrates thinks that) all our desires, indeed, all our mental states, refer to real, rather than intentional objects. So when Socrates says ‘You didn’t want to do action x’ it is because the candidate object of your wanting or not wanting, ‘action x’, must be the action you actually did, along with all its actual (bad) consequences. My ‘Plato on the Attribution of Conative Attitudes’, *Archiv für Geschichte der Philosophie*, 88 (2006) [= ‘Attribution’], 127–62 contains extensive discussion and criticism of Penner’s interpretation which I won’t repeat here. But the key textual difficulty this interpretation faces from the *Gorgias* is that Socrates says that there are desires which are in need of disciplining (491d–e, 493d–494a), the fulfilment of which brings pleasure, which may be good or bad. But why should desires that can only obtain if their objects are good need to be disciplined? Penner says that the *Gorgias* contains non-Socratic ‘elements’ (150 n. 3), and proposes that desires for things that are not good may exist as hankerings, itches, and drives, which cannot motivate a particular action (201f. n. 45). Penner must think the reason to discipline such desires is their effect on our actions via an effect on our beliefs.
31. K. McDaniel and B. Bradley, ‘Desires’, *Mind*, 117 (2008), 267–302.

32. Cf. Seneca, *On Peace of Mind* 13.2–14.2 in J. W. Basore (trans.), *Seneca Moral Essays*, vol. II (London, 1932).
33. McTighe, ‘Socrates on Desire’, raises this objection (205–7) and argues that elsewhere in Plato, *boulesthai* is for the apparent good, and that what Plato really believes is that the good exerts a ‘teleological lure’ on us, not that it is the object of our psychological attitudes. So his explanation of the present argument is that Socrates is effecting a ‘purgation of [Polus] . . . conceit’, for Socrates ‘has no reason necessarily to adhere to the standards of logic or fair play—and possibly reason not to.’ (226) I agree with McTighe about the teleological lure, but not that Socrates’ argument in the *Gorgias* is (deliberately) fallacious or that elsewhere in Plato *boulesthai* is for the apparent good.
34. Exemplifying the class of ‘goods’ by such ‘popular’ goods, McTighe argues, tricks Polus into accepting Socrates’ (3) as the claim that we do intermediate actions for the sake of ends we believe to be good (‘Socrates on Desire’, 206–7). Attempting to defend Socrates, Wolfsdorf argues that the examples are ‘dialectical’, not endorsed by Socrates but asserted for the sake of simplicity, given his argument’s focus on instrumental desires (‘Rhetoric’s Inadequate Means’, 124, 127). Wolfsdorf is right about the argumentative focus of this passage, but Plato is writing a dialogue, not a passage, and Socrates’ assertion that health and wealth are goods in this passage had better be consistent with his claim, later, that the whole of happiness is ‘in’ education and justice (470e). Of course, Socrates need not explain all his views about goodness in every dialogue, but Plato ought to have in mind some way in which Socrates’ claims are consistent, if he is not to put him on all fours with the likes of Polus and Callicles.
35. My appeal to other dialogues for parallels to the objective reading of Socrates’ discussion with Polus in the *Gorgias* is not meant to be of the form, ‘Plato believes P about desire for the good, therefore we are entitled to assume Socrates says/thinks P about desire for the good in the *Gorgias*’. Rather, the parallels are meant to show that Plato takes P about desire for the good as widely shared and so as common ground in discussion. That we desire really and not only conventionally good things is not the conclusion of philosophical argument but a (frequently used) starting point.
36. An anonymous reader objects that the desire requirement for explanation is a Humean prejudice, but I am not arguing that reason cannot move one without desire; rather, I am explaining why appearances of goodness move us.
37. One may think that instrumental desire is instead dependent on what we know/believe/are means (or not), tracking belief rather than truth. But this is only plausible if we think belief tracks the truth, and I do not see that we have any more reason to think this of belief than of instrumental desire/desire for good things.

38. E. Dodds (ed.), *Gorgias* (Oxford, 1959) (= *Gorgias*], 236: the 'inner man . . . who is an immortal rational being', distinct from 'the empirical self', is the true subject of wanting.
39. N. Gulley, 'The Interpretation of "No-One Does Wrong Willingly" in Plato's Dialogues', *Phronesis*, 10 (1965), 82–96.
40. Segvic, 'No-One Errs Willingly', 10.
41. As argued in Kamtekar, 'Attribution'.
42. Aristotle rejects the *Gorgias* requirement that the object of wanting be really good on the following grounds: it will make the person who chooses incorrectly want what is not an object of wanting—but presumably the agent's choice, pursuit, etc. are reasons to say she does want the object, for they would have been reasons to say she wanted the object if her judgement as to its goodness had been correct. But, Aristotle acknowledges, saying that the object of wanting is the apparent good has the consequence that there is no natural object of wanting (this, I suppose, is what a creature's natural desires are for). His solution is to make the natural object of wanting what the good person wants.
43. Dodds, *Gorgias*, 236. One might think to defend Socrates on the grounds that he only *deattributes* desires, but Socrates says of tyrants and orators that they don't do any of the things they want (*ouden . . . poiein hôn boulontai*, 466d8–e1), which certainly seems to be a positive attribution of desires that are likely not for the things that seem good to the tyrants and orators.
44. G. E. M. Anscombe, *Intention* (Ithaca, 1957) [= *Intention*] [s. 36], 67.
45. Against Callicles' image of a naturally powerful man (483c), with Socrates asking 'powerful (better, superior) in what?' until Callicles settles on 'intelligent in the affairs of the city and brave' (491c), whereupon Socrates asks if self-control/mastery over one's appetites isn't required for power too.
46. Agnes Callard has brought to my attention one passage in the *Gorgias* that countenances a situation in which doing bad might be the means to one's end: if one ought to do badly by/harm (*dei tina kakôs poiein*) one's enemy, one ought to—carefully, lest one be unjust to oneself/suffer injustice (*adikêtai*)—keep him from being brought to justice for any wrong he has done (480e–481a). I suppose the risk to oneself is that in doing badly even by one's enemy, one might become unjust oneself.
47. I consider this to be a dialectical premise: 'pleasant or good/beneficial' is Polus' (and perhaps a common naturalizing) analysis of 'fine'/'admirable'.
48. Santas, 'Socratic Paradoxes', 162, endorses this consequence.
49. Vlastos, 'The Socratic Elenchus'. T. Brickhouse and N. Smith, 'Socrates' Elenctic Psychology', *Synthese*, 92 (1992), 63–82, propose that both the attributions of beliefs and wants are attributions to the 'true self' which elenchus helps us to come to know.

50. In the Callicles section of the dialogue, Socrates argues that although Callicles despises the people, he is bound to agree with them, although not himself (481d–482c). I argue that this is on the grounds that Callicles has inherited their conception of the good in ‘The Profession of Friendship’.
51. Thanks to an anonymous referee for this observation.
52. Cf. *Sophist* 228c–d: no soul is willingly ignorant; ignorance arises when the soul reaches after truth but misses; *Laws* 864b: belief strives after truth. Thanks to an anonymous reader for these two references.
53. J. Moss, ‘Pleasure and Illusion in Plato’, *Philosophy and Phenomenological Research*, 72 (2006), 503–35, takes the view that pleasure is the apparent good. Moss has an easier time getting to this conclusion than I do because she takes the *Gorgias* wanting argument to say that we desire what seems good, and adds to this the plausible thought that pleasure seems good (512). But this ignores the very important distinction between wanting (*boulesthai*) and appetitively desiring (*epithumein*).
54. For Brickhouse and Smith, *Socratic Moral Psychology*, appetite is a paradigmatic non-rational motivation that affects action indirectly by influencing beliefs about good and bad: appetite presents its object to belief as good or bad (52 n. 6), and the resulting belief about good or bad combines with the desire for good to produce an action.
55. B. Holmes suggests that the *Gorgias* thinks about the soul by analogy with medicine’s thinking about the body, concluding that just as bodies have an internal principle inclining them to health but also tendencies that would lead to disease unless ordered by the expertises of medicine and gymnastics, so too, souls have an internal principle inclining them to the good but also appetites that would lead to psychic disease unless disciplined (by justice and legislation). See ‘Body, Soul, and Medical Analogy in Plato’, in K. Bassi and J. P. Euben, *When Worlds Elide: Classics, Politics, Culture* (Lanham, 2010), 345–85. Holmes supposes that the appetites require disciplining because they are insatiable, but Socrates seems to say that only the appetites of fools are insatiable (one’s jars may be leaky or sound); the problem with the appetites seems rather to be that they are (like the *Protagoras*’ appearances) inferior guides to goodness, and for that reason should be subordinated to something that tracks the truth about goodness.
56. An anonymous referee objects: doesn’t this also give Socrates two kinds of people after all, those who desire only good things and those who desire good things and bad things believing they are good? The answer is that Socrates does want to end up with two kinds of people, the virtuous and the vicious, but the distinction between them isn’t ultimately one of desire—the kind that also desire bad things believing they are good are distinguished by what they believe is good.

57. Santas, 'Socratic Paradoxes', 156: 'what Socrates is denying is that bad things are the intended objects of these people's desires'; cf. Devereux, 'Socrates' Kantian Conception of Virtue', 401–2.
58. Santas, 'Socratic Paradoxes', understands Socrates to be saying, 'no one desires to be miserable (in the sense that to be miserable is never the intended object of anyone's desire)' (157), when Socrates says no one wants (*bouletai*) to be miserable (78a4). But on Santas' own account this should be a claim about the actual and not only the intended object (152 n. 15).
59. See T. Penner and C. Rowe, 'Desire for Good: Is the *Meno* Inconsistent with the *Gorgias*?' *Phronesis*, 39 (1994) [= 'Desire for Good'], 1–25. I discuss their interpretation in detail in 'Attribution' and so will not repeat that discussion here.
60. A. Callard, 'Everyone Desires the Good: Socrates' Protreptic Theory', *Review of Metaphysics*, 70 (2017), 617–44.
61. What Santas, 'Socratic Paradoxes', gets importantly right (and I did not give him credit for this in 'Attribution') is that Socrates' point at 77d–e is that to say 'he desires bad things' of a man who believes the things he desires are good when they are in fact bad is 'misdescribing the object of his desire' (154). But Santas does not explain why this matters to Socrates—and I think he doesn't see why it matters, because he says 'the only relevant thing that can follow from this [viz. 'the statement that they thought the things . . . they desire were good things'] is that good things are the *intended* object of their desires' (156). But Socrates says another, more relevant, thing that follows: if they desire things they think are good, they desire good things.
62. Because the protasis of this conditional is in the present tense, it is strictly speaking not contrary-to-fact (viz. 'if there were such a person'). For parallels in which a present tense in the protasis is followed by an optative in the apodosis, see *Apology* 17b: 'If they mean (*legousin*) this [viz. that an orator is one who speaks the truth], I would agree (*homologoiên*) that I am an orator'; Sophocles *Philoctetes* 116: 'if this is how things are, then it would be necessary to track them down (*thêrate' oun gignoit' an*). Seth Schein (personal communication) describes the latter's use of present tense in the protasis and optative in the apodosis as indicating a mental struggle to accept the shocking consequence of a reality; perhaps the *Hippias Minor* conditional should be understood on this model.
63. Cf. Aristotle's comment that the liar is not merely the one who has the power to lie, but the one who has an inclination to lie, for its own sake/under that description (*Metaphysics* 1025a1–13).
64. What about the person who runs fast unwillingly, e.g. not because training has made him strong and fast, but out of fear?



65. So I disagree with N. Gulley, *The Philosophy of Socrates* (New York, 1968): 'ability means ability to achieve one's aim, to succeed in what one wants to do . . .'; 'knowledge here means expert skill or competence in achieving a given aim and not knowledge of what the aim is' (85); 'ability is a necessary but not a sufficient condition of achieving the "right" aim . . . Here, Socrates thinks, the analogy between moral knowledge and the knowledge belonging to professional skills breaks down' (86). I don't think Socrates is so quick to reject the skill analogy and I think there is a great deal to the analogy.
66. The breakdown is reminiscent of Gorgias' *Encomium* 8–13, except that in that text these effects are all brought about by speech.
67. Zeyl even translates *akousiôtata* 'entirely beyond our control'.
68. In his commentary on the *Timaeus*, F. Cornford relates this to the *Gorgias*' distinction between the true/rational self that is the subject of wish (*boulêsis*) and the appetites ((trans. and comm.), *Plato's Cosmology* (New York, 1952), 348). There are two differences to note: first, unlike the *Gorgias*, the *Timaeus* does not single out *boulêsis* as the conative attitude contrariety to which determines unwillingness (although presumably it is still contrariety to our orientation to our real good that makes something unwilling), and second, *Timaeus* adds to the list of causes of being in a state unwillingly or acting unwillingly the causes external to the agent to which his character can be traced. A precedent for this might be Protagoras' claim that our characters are the product of our god-given nature (*Protagoras* 322c–d) and the education we receive by growing up in cities (which makes people good) or not (which makes them very bad) (327c–e)—although Protagoras himself does not comment on the implications of his account of character-formation on the voluntariness or involuntariness of actions, and considers punishment only as a deterrent (324a6–b5).
69. In *Plato on Punishment* (Berkeley, 1981), M. M. MacKenzie says the ignorance in question is wrongdoers' ignorance of their own psychological defectiveness, because what makes wrongdoing unwilling is wrongdoers' not knowing what virtue and vice are. (She is followed in this by C. Gill, 'The Body's Fault? Plato's *Timaeus* on Psychic Illness', in M. R. Wright (ed.) *Reason and Necessity: Essays on Plato's Timaeus* (London, 2000), 59–84.) But this makes rightdoing unwilling as well, since not even Socrates knows what virtue and vice are. Further, why couldn't a person know perfectly well that what's wrong with her is that she lacks virtue, and watch, helplessly, as she does wrong? (That, famously, is Alcibiades' condition.)
70. M. Schofield, 'Injury, Injustice, and the Involuntary in the *Laws*', in R. Kamtekar (ed.), *Virtue and Happiness: Essays in Honour of Julia Annas*, *Oxford Studies in Ancient Philosophy Supplementary Volume* (Oxford, 2012), 103–14 at 109. Schofield argues the Athenian upholds the 'Socratic paradox' that no one does wrong willingly but without relying on 'Socratic

intellectualism', viz. the view that no one does bad or wrong knowingly (106). For in the Athenian's claim that ignorance (*amathia*) is the cause of wrongdoing (863c), 'ignorance' includes not only the absence of some information, but also 'losing one's grip on [knowledge or reason]', as in 'folly' in the *Timaeus*, and in the claim that hating something you judge good and fine is ignorance (689a, 105). What Schofield claims to be the case in the *Laws* I claim to be the case throughout the corpus: 'no one does wrong willingly' never relies on 'no one does wrong knowingly', although ignorance is of course one explanation for one's acting contrary to one's wish.

71. A similar line is taken by T. Saunders in 'The Socratic Paradox in Plato's *Laws*', *Hermes*, 96 (1968), 421–34, which argues that the Athenian wants to say that the same act, considered as a harm, is willing if it was open to the agent to do or not do it, but considered as an injustice, is unwilling as per 'no one does wrong willingly' (424). Saunders concludes that according to Plato, we may justly punish a man for doing what he 'wanted' to do even though he did not 'really' want to do it but was overcome by emotion or misdirected by ignorance, because Platonic punishment is corrective (434).
72. Oddly, Schofield says that the Athenian's argument is designed to eliminate 'wrongful injury' (111), but the Athenian's argument is designed to eliminate 'involuntary injustice', not 'wrongful injury'; indeed, the Athenian legislates with respect to wrongful injury (862d, cf. (L9)). My account of the two distinctions is not so far from J. Roberts, 'Plato on the Causes of Wrongdoing in the *Laws*', *Ancient Philosophy*, 7 (1987), 23–37, except that Roberts thinks what makes injustice unwilling is 'the psychological impossibility of wanting to be defective', which she equates with the 'Socratic paradox' from the *Protagoras* (31).
73. S. Sauvé Meyer, *Aristotle on Moral Responsibility: Character and Cause* (Oxford, 2011), 132.
74. S. Bobzien, 'Choice and Moral Responsibility (NE iii.1–5)', in R. Polansky (ed.), *The Cambridge Companion to Aristotle's Nicomachean Ethics* (New York, 2014), 81–109 at 98.
75. Aristotle explains that someone who fails to return a deposit because he is compelled, and unwillingly (*akôn*), does not do injustice (*adikein*) except coincidentally, and someone who returns a deposit unwillingly and out of fear does justice (*dikaia prattein/diakiopragein*, the contrary of *adikein*) only coincidentally. This seems to be in some tension with NE III.1's characterizations of such actions as 'mixed' and 'more like the voluntary'.
76. Bjorn Wastvedt has raised the excellent question whether Aristotle is right to assume that everyone knows that doing vicious actions will make one vicious, rather than incontinent. For it's one thing to know that overeating, for example, will make me such as to overeat again, but quite another to know that it will eventually make me not even care that I'm overeating, and

to give up trying not to. Perhaps Aristotle only thinks everyone knows the former—that is, after all, the analogy between virtue and skill, whereas the influence of our actions on what appears good to us is different from the skills, and one would have to be very deliberately shaping one's own character to be thinking ahead about how what one does will influence what appears good to one's future self.

77. But cf. *Laws* X 904c *boulêsesin*.

# 4

## The Divided Soul

### 4.1 Introduction

Let us review our results. Chapter 2 has argued—contrary to the mainstream interpretation—that rather than taking psychological eudaemonism, the view that we always do what we believe is the best of the things we can do, to be the foundation of a Socratic intellectualist moral psychology, the *Protagoras* explores it as an alternative to treating actions as the outcome of competing forces in the soul. And Chapter 3 has argued that the *Gorgias* and many other dialogues posit our desire for the real good (rather than for what we believe is good) as the basis for claiming that self-destructive and wrongful actions are unwilling, with the *Gorgias* claiming in addition that we do not want to do such actions even if we judge them best, and the *Meno* claiming that we do not desire things we believe are good unless they are also actually good.

But if this is right, then why does Plato divide the soul with such fanfare in the *Republic*? The mainstream interpretation has a ready answer: in the early dialogues Socrates believes that an agent's every action is motivated by her judgement that this action is better than the other actions she can do; in the *Republic*, Plato repudiates that belief and admits that our souls have multiple sources of motivation, some of them good-independent, each of which can motivate us to act on its own, even contrary to a judgement about what is best to do. A minority interpretation maintains that the *Republic* does not break with the psychological eudaemonism of the early dialogues but elaborates it by defining good-independent motivations that influence our judgements about what is best to do.

Any adequate interpretation of this dialogue's claims that our souls are divided into parts capable of certain sorts of conflict, including conflict with reasoning (436b–441c, 602c–604d), and its identification of these parts by their characteristic objects of pursuit—knowledge, honour, and

pleasure (580d–581c)—must be consistent with its statement that every soul pursues the good and does everything for its sake (505d–e). The present chapter agrees with the majority interpretation that the *Republic* breaks with the *Protagoras* by admitting the existence and independent motivational efficacy of *multiple* motivational sources, but disagrees with that interpretation that any of them are good-independent. The point of soul-division, I argue, is to admit that all these motivations are *ours*, even when acting on them undermines our pursuit of our own good (so that, as we saw in Chapter 3, we act ‘unwillingly’). Thus what matters in Plato’s arguments for soul-division is to establish that there are in our soul multiple, independent, and potentially conflicting sources of motivation, all of them equally part of our nature (Section 4.2). Plato sometimes treats these sources of motivation as themselves agent-like subjects of beliefs and desires, each capable of generating action on its own, and also gives them personae, as pursuers of the good under some conception of the good (wisdom, honour, or bodily pleasure), for two reasons. First, doing so enables him to give psychological eudaemonist explanations of each of our actions while allowing that we sometimes act contrary to our reasoned judgements, and downstream from this, to explain how we come to have the characters we do, using the political model of a faction taking power and establishing a new constitution. Second, it provides a useful way of evaluating our psychological attitudes when we are not in a position to assess their contents, for example on some occasions when they come into conflict with one another (Section 4.3). The good-directedness of each of the soul-parts is confirmed by the teleological account of the soul in the *Timaeus* (Section 4.4).

## 4.2 Soul-Division in *Republic* IV and X

In *Republic* IV, Socrates raises the difficult question of whether, when we do the different sorts of things we do—desire bodily pleasures, grow angry, or learn—we do them by the agency of one and the same thing (*tôî autôî*) or three different things (*trisin ousin allo allôî*, 436a9). This is the same question that he asks in the *Phaedrus*: is he ‘more complicated than Typhon’ (*poluplokôteron Tuphônôs*) or something ‘simpler’ (*haplouteron*) (230a4–5)? In the *Phaedrus*, rather than argue that we are multiple, he assumes it as ‘likely’ (246a5), and later describes the method by which one would establish how many: one would need to apply to

the soul the method of Hippocrates, which begins by finding out whether the thing being investigated is simple (*haploun*) or multiform (*polueides*), and goes on to determine, for each simple, what is its power (*dunamis*) of acting and being acted upon, and in relation to what object (*pros ti*) (270d). Here in the *Republic* he actually argues that the soul is multiform, applying a principle of opposites to cases of conflict to demonstrate that the soul is a ‘many’. According to this principle:

The same thing (*t’auton*) won’t do or suffer opposites, at the same time (*hama*) in the same respect (*kata t’auton*),<sup>1</sup> in relation to the same thing (*pros t’auton*); if these things happen [viz. the doing or suffering of opposites in the same respect in relation to the same thing at the same time], we’ll know it is not [one and] the same thing (*t’auton*) but many (*pleiô*). (436b8–c1, cf. 436e–437a)

But one or many *what?* The terms Socrates uses for the elements in the soul he is investigating are in general metaphysically non-committal, for example ‘kind’ (*eidos*), ‘part’ (*morion*), ‘the thing bidding them’ to drink (*to keleuon*), ‘the thing forbidding’ (*to kôluon*). His opening question, ‘with/by the agency of what do we desire bodily pleasures, or grow angry, or learn?’, indicates that he is asking how many sources of psychological activities or movements are in us.<sup>2</sup>

Why does it matter whether we are one or many? Attention to the place of soul-division in the dialectic of the *Republic* yields an answer closely connected to the issues we have been considering in previous chapters. As we saw in Chapter 1, Socrates’ task in the *Republic* is to give an account and defence of individual justice, and he does so by analogizing the individual to a (partite) city. But we are now in a position to see that the alternative account of and attack on justice against which Plato frames Socrates’ task makes a very specific assumption about human nature. In Book II when Glaucon is representing the amoralist challenge to justice, he says that to have more and more, and more than others, is what every nature naturally pursues as good, but this nature is led to honour equality by law and force (*tên pleonexian, ho pasa phusis diôkein pepuhkan hôs agathon, nomôi de biai paragetai epi tên tou isou timên*, 359c4–6)—where equality is the compromise reached by people who agree not to harm others in order to avoid being harmed themselves (358e–359c). Glaucon’s challenge supposes that human nature is not many, but one, oriented towards having more—more than one has, and more than others. Justice and law, on this picture, is a non-natural,

external, force that checks our natural desire for our good.<sup>3</sup> While scholars have rightly commented on the difference in content between what the amorlists think is good—the ‘more and more’ of appetitive satisfaction—and what Socrates argues is good—a reason-ruled life—Socrates and the amorlists also differ over whether human nature is simple or complex. The main argument of the *Republic* requires Socrates to correct their too-simple conception of human nature and, as a consequence, of human good.

Already in Book II Plato begins to show the poverty of this picture of the relationship between nature and law. Glaucon himself recognizes that the life of necessary appetitive fulfilment described in Socrates’ first, ‘healthy’, city (372a–d) falls short of being a good life for human beings (thus, he objects that it is a city for pigs), and it would seem to be law or custom (*nomos*) that is responsible for some of the particular desires and desire-fulfillments that make human lives good. The couches and tables on which Glaucon would have citizens dine, for example, are the furniture of that elite cultural institution, the symposium, which would not come about without custom and education.<sup>4</sup> And when Socrates agrees to set up the city so citizens can enjoy the goods Glaucon thinks happy humans must enjoy, he says that such a city requires guardians, who must be educated lest they themselves destroy it (373a & ff.), and that the education they receive purifies the city (399e). But the education they receive is determined by the law that aims at the good of the whole. These points suggest that law or custom is needed to direct our good-seeking desires so that they can achieve the good which they seek; in our pursuit of the good, law completes human nature rather than opposing it. But until Glaucon and readers of the *Republic* see that our natural desires have many independent sources, both bodily and non-bodily, it’s difficult to see law, and justice, in this new role.

To argue for the soul being many in *Republic* IV, Socrates, after introducing the principle of opposites, lists several psychological acts that have opposites: assenting and denying, desiring/going after (*epihesthai*) and rejecting (*aparneisthai*), taking for oneself (*prosagesthai*) and pushing away (*apotheisthai*) (437b1–4). Included in this class, he says, are thirst, hunger, and all the appetites (*holôs tas epithumias*), being willing (*ethelein*), and wishing (*boulesthai*). (Perhaps psychological states like being appeared to, or imagining, are excluded from the class as not having opposites.) Socrates then gives a further characterization of having an appetite:

The soul of one who has an appetite (*tou epithumountos*) goes after (*ephiesthai*) what he has an appetite for, or<sup>5</sup> brings to himself (*prosagessthai*) whatever it is he wishes (*boulesthai*) to come to be his, or insofar as he is willing (*ethelein*) for something to be given to him, nods in assent to this as if to a questioner.

(437b7–c6)

The person who has an appetite for something might pursue it, or wish it to be his and bring it to himself, or be willing for someone to give it to him and so accept it. Socrates shouldn't be seen here as assigning these three attitudes to different parts of the soul, for he has yet to show that the soul is divided. Rather, he seems to be characterizing our psychological attitudes grossly and extensionally, presumably for the reason that if he specifies desires too narrowly (e.g. 'appetitive desire' or 'rational desire') or in terms of the description under which things are desired (as he does at *Charmides* 167d–e, *boulêsis* being for the good and *epithumia* for the pleasant), he is not going to be able to show that we are subject to oppositions that are simultaneous, with respect to the same thing, and in relation to the same thing, so as to generate a 'many'. (Later, when he has established that there are in our souls many sources of motivation, it will be useful to think about the description under which these desire or pursue their objects.)

To introduce thirst, the desire that generates the first conflict to separate out the reasoning and appetitive elements in the soul, Socrates begins with a definition: thirst, insofar as it is thirst, is an appetite for drink as such. Thirst (by itself) is not for cold drink, which would be the object of thirst qualified by the presence of heat. Conversely, if it were much thirst then its object too would be qualified as not just drink, but much drink (437d–e). This claim about thirst is based on a principle of correlatives which Socrates illustrates with a number of examples (438b–e): the greater is such as to be greater than the lesser; the going-to-be-larger is such as to be larger than the going-to-be-smaller; knowledge is of what can be learned—and knowledge of building is different from the other kinds of knowledge. In these cases the relationship between the two correlatives seems to hold uniquely, invariably, and transparently.<sup>6</sup>

Having defined thirst, insofar as it is thirst, by its correlative object, Socrates draws a lesson about desire-attribution:

...let no one clamour at us being uninvestigative, [saying] that (*hôs*) no one desires (*epithumei*) drink but rather good drink, nor food but good food, for (*gar*)



everyone desires good things, so that (*oun*) if thirst is a desire, it will be a desire for good drink or whatever, and similarly with the others. (438a)

Many commentators say that this passage introduces good-independent desires using the example of thirst: thirst is not desire for drink as good, for if it were, it would be desire for the good, not for drink.<sup>7</sup> A good-independent desire is taken to be a desire for something not under the guise of, or thought to be, or conceived as, good. The idea is that in this passage, Socrates asserts that there are some desires, for example bodily desires like thirst and hunger, that are simply desires for drink or food, not drink or food conceived as good. According to an alternative, minority interpretation, this passage only says that although thirst is in fact always desire for drink as good, the *definition* of thirst is 'desire for drink'—'good' doesn't distinguish it from other desires, which are all for something or the other as good.<sup>8</sup>

Both these interpretations import the idea of desiring as good into the text. For all Socrates says is: don't infer, from everyone's desire for good things, that they desire only the good instances of any given kind. Rather, the thirsty person, insofar as he is thirsty, wants drink, not only good drink—even though he desires what's good. This implies that we can't infer, from the badness of a particular drink, that the thirsty person doesn't desire it—even though it is true that the thirsty person desires good things and this drink isn't good. In this passage, at any rate, Socrates is neither denying that we desire whatever we desire thinking that it is good, in order to introduce good-independent desires, nor affirming that we desire whatever we desire thinking that it is good, while denying that it is a basis on which to define or distinguish desires. He is simply silent about the description under which something is desired, or about whether the subject of desire must believe that the object of desire is good. Instead, he preemptively blocks the following objection to the conflict scenarios he is about to provide: the thirsty person who doesn't drink on the grounds that the drink is bad doesn't desire the drink he doesn't take, because what he desires is good drink (which we can deduce from his natural desire for good things), and this drink is bad!<sup>9</sup> Blocking this objection is crucial for his argument for soul-division, because the argument depends on attributing conflicting desires to the same individual, and if some these can be deattributed on the grounds that they conflict with the desire for the good, soul-division will never get off the ground.

We saw in Chapter 3 that in the *Gorgias* and *Meno*, Socrates deattributes false or harmful psychological attitudes on the grounds that they conflict with our natural desire for good things. There, I tried to motivate Socrates' inference both by showing its roots in the sophistic opposition between our natural pursuit of our own advantage and compelled actions contrary to it, and by arguing for its plausibility in the case of instrumental desires. But the inference is hardly unproblematic: first, it seems arbitrary that behaviour indicative of desiring—pursuit, acceptance, and so on (437b)—should always be trumped by contradiction with our actual good. Second, if we don't desire the things we pursue or accept, what explains our pursuing or accepting them?<sup>10</sup> Perhaps it is reasonable to infer from a desire for things that seem good, on the grounds that they seem good, to a desire for things that are really good. But the converse inference—that one doesn't desire things that seem but are not in fact good—is not reasonable (unless one's desire is an instrumental desire and so is closely governed by the truth about the instrumentality of the object desired).

As noted in Chapter 3, the appetites are a clear case of non-instrumental desire. We do not feel hungry and thirsty on the grounds that eating or drinking is our means to securing some further end of ours, even though our having these sorts of desires may be instrumental to our health. Of course we *may* want to eat or drink as a means to securing some such further end. For example, even though I am not hungry when I wake in the morning, I might want to eat breakfast on the grounds that a good breakfast is essential to good health (or productive mornings, or something else). But this is not the same as being hungry. Conversely, even if it serves a living creature's good to be hungry and to be moved by hunger to pursue food, and even if these are the means by which it nourishes itself and lives out its characteristic life, the creature need not have a representation of its hunger for and pursuit of food as instrumental to staying alive or living a certain kind of life, and it need not fulfil its desire for food for that reason. Indeed, natural teleology provides a consideration against deattributing particular desires simply on the grounds that their fulfilment would be contrary to our good or happiness, for even if *having* desires responsive to certain sorts of appearances and with a certain range of fulfilments serves our good, these are not guaranteed to get us the good in every circumstance (I discuss this further in Section 4.4). Admitting that we can desire something actually bad is not evidence that not all

desires are good-directed, but instead evidence that good-directed desires can fail to hit the target—just as our rational capacities can fail to hit upon the truth despite their aiming to or being designed to hit it. Of course, none of these thoughts about teleology are in our passage; but neither are thoughts about the description under which we desire things. Instead, the passage just blocks a deattributing inference.<sup>11</sup>

After these preliminaries about opposites, correlatives, qualification, and attribution, we are at last ready turn to the case for *Republic IV*'s first division of the soul, into an appetitive and a reasoning part:

- (D1) The soul of the thirsty person, insofar as he is thirsty, wishes (*bouletai*) nothing other than to drink, and wants (*oregetai*) it and is impelled (*hormai*) to it (439a–b).
- (D2) So (by the principle of opposites) if something draws the soul back, it must be something different in the soul (than what impels it).
- (D3) Sometimes thirsty people don't want (*ethelein*) to drink: one thing in their soul bids them to drink, as a result of a bodily condition, and another thing forbids them, as a result of calculation.
- (D4) Since the bidding is the result of a bodily condition and the forbidding is a result of calculation, the two principles in (D3) are the appetitive and rational parts of the soul (439c–d).

(D1) echoes one of the preliminaries we have just gone through (437c): thirst (an appetite) gives rise to both a psychological attitude directed to an action (*bouletai*) that will procure the object desired (*oregetai*) as well as a motion (*hormai*; I'll say more about the psychological and kinetic characterizations in Section 4.3). So there is in the soul of the thirsty person who doesn't drink both a pro-attitude towards something (the drink, getting drink, drinking), and a con-attitude towards the same thing.<sup>12</sup> (D3) then stipulates that in the example, the thirst arises from a bodily condition and the forbidding to drink from calculation. Once the conflict is established, Socrates' stipulation identifies the sources of the two conflicting desires. He then generalizes: in such cases, the forbidding (*to kôluon*) arises from calculation, and the leading to it from bodily conditions and diseases (439d). This would be question-begging if he took it to establish a whole lot about the character of the

parts, but saying they are *logistikon* and *epithumêtikon* is naming the parts by what they were doing in the example.<sup>13</sup>

One might wonder why the principle of opposites gives Socrates only three soul-parts. We seem to be able to come up with conflicts that would divide some of the soul-parts once again. For example, in the case of appetite, it seems that when I have a stomach bug I can be hungry for and nauseated by the same food. But perhaps Plato does not think the appetitive part is unified: Socrates later says that the oligarch's appetite for wealth conflicts with his other, spendthrift appetites (554b–e).<sup>14</sup> At the end of Book IV, Socrates says that the just person harmonizes the three soul-parts 'and if there happen to be others in between' (443d7). And in *Republic X*, Socrates never calls the inferior part of the soul 'appetitive' (*epithumêtikon*) or identifies its motivations as bodily but instead uses periphrastic expressions that name it by the various reason-opposed responses in the cases he considers: 'the thing that leads to recalling sufferings and to lamenting and is insatiable' (*to de pros tas anamnêseis te tou pathous kai pros tous odurmous agon kai aplêstôs echon* 604d8–9) 'the lamenting thing' (*to thrênodon* 606a8–b1), 'the pitying thing' (*to eleinon* 606b7–8), and 'the thing forcibly restrained in private misfortunes' (*to biai katechomenon tote en tais oikeiais sumphorais* 606a3–4). It is worth bearing in mind that Socrates warns at the outset of the argument that he will not be able to give a precise answer to the question whether the soul has three parts like the three parts of the city using the methods of argument he is using, but will only be able to give an account adequate to establishing what justice is in the soul (435c–d, cf. 504b). Still, there is no question that reasoning, spirited, and appetitive motivations are the important ones to distinguish, at least for the purposes of understanding justice.

Socrates cannot establish any more than a plurality in the soul by the principle of opposites alone, for every new conflict, by itself, only establishes another plurality whose connection to previously identified elements in the soul is yet to be determined however, it is always in principle possible to assign the conflicting attitudes to previously established parts. Enumerating the parts, distinguishing the plurality in a conflict from previously established parts, or identifying them with these, requires appeal to the power of or function performed by the parts. Consider Socrates' examples of the spirited part's operations: first, Leontius, desiring to look at some executed corpses and simultaneously (*hama*) feeling

disgust at himself and turning himself away, after some struggle is overcome (*kratoumenos*) by his desire and rushes towards the corpses, berating his eyes: 'Look, unhappy ones, fill yourselves with the beautiful spectacle!' (439e–440a). Why is this not reason's response but that of a third part with which we anger? Glaucon observes that small children, who lack reasoning, also grow angry; Socrates adds that other animals do too (441a–b). But are all these reactions from the same source in the soul? A second case of conflict is supposed to be of the reasoning part against the angering: Odysseus, home at long last in Ithaca, hears his maidservants on their way to meet the men who have been trying to take his place, but restrains his angry desire to kill them at once because he has calculated (*analogisamenon*) that doing so will expose him to discovery and danger (441b, cf. *Odyssey* 20.17–18). To establish that Odysseus' anger belongs to the same kind or part as led Leontius to berate himself—rather than belonging to appetite or some further part—we need to establish that the two responses are instances of the same function, so far just described as 'being angry'. The fact that appetite's source is given as bodily conditions and reason's source as calculation is crucial supplementation to the 'many' established by applying the principle of opposites to cases of psychic conflict. Unless we can identify a corresponding source for the spirited part of the soul (society's values? the person's self-esteem?) it is missing one support for being a third kind or part in this argument. Later, Socrates identifies the three soul-parts by their characteristic objects of desire and pleasure, namely wisdom, honour, and appetitive gratification (*Republic* 580d–581c). So perhaps it would establish that Leontius' and Odysseus' responses arise from the same soul-part to show that they both arise out of a concern for honour. (Chapter 5 is devoted to why Plato argues for a third, spirited part of the soul, so I will leave this question till then.)

I have been arguing that the point of *Republic* IV soul-division is to establish a plurality of independent and potentially conflicting sources of motivation in the soul, a requirement for which is that a motivation that detracts from one's pursuit of one's good not be deattributed on that ground. This gives Socrates an account of a complex human nature that, contrary to the amoralist sophists, is not simply compelled by law, for law enables the better parts of it to manage the worse for the sake of the complex whole. One alternative is to suppose that by including good-independent desires among our motivations Plato is being more realistic

than the intellectualist Socrates of the early dialogues. Terry Irwin proposes that soul-division responds to a 'demand for explanation'.<sup>15</sup> The desire for good can't explain my desiring and choosing to drink when my reasoning about what is better has concluded that it's better to abstain. So in reply to the Socrates of the *Protagoras*, according to whom 'we explain and understand an agent's action only if we refer to some end, and ultimately to some self-explanatory end [viz. happiness]', so that to say someone acts contrary to his reasoned choice because of a non-rational passion like fear is to render his action unintelligible (209), in the *Republic* Plato claims that non-rational passions do render the actions they cause intelligible insofar as they belong to a kind, 'appetite', typified by 'basic biological urges or drives that we share with other animals' (210). In other words, Plato realizes that psychological eudaemonism leaves akratic actions and contrary-to-reason feelings unexplained, and so posits appetitive desire as a new explanans for them; this explanans is not ad hoc, for it also explains other actions unexplained by psychological eudaemonism, namely the actions of non-rational animals.

We can certainly see positing the class of appetites as rendering certain actions intelligible in a new way. Indeed, collecting motivations into a kind or class, and delimiting that class by an account of the origin of these motivations in bodily conditions, looks like the exercise in collection and division en route to definition which Plato uses in later dialogues like *Statesman*, *Sophist*, *Philebus*, and *Phaedrus*. But why is this alternative preferable to the *Protagoras*' redescription of the passions as appearances of, or appearance-based beliefs about, goodness? If Plato's purpose were to improve on the psychological explanations promised by the *Protagoras*, it seems he would need to answer this question. Irwin suggests that all he needs to do is to give an account of akratic action that is not incoherent as was the Many's account in the *Protagoras* (p. 209), but that would seem only to put psychological eudaemonism and the divided soul on all fours. Further, parts of the soul which are good-independent in their pursuits would conflict with Socrates saying, 'every soul pursues the good in everything it does, and does everything for its sake' (*Republic* 505d–e), for this would seem to include the souls of non-human animals and plants, yet their souls seem to be appetitive and in some cases spirited, but not rational. (I treat the topic of psychological explanation more fully in Chapter 6.)

By contrast, on my account of the point of soul-division—to establish multiple sources of motivation in the soul as natural and as belonging to us—the correction being made is not to psychological eudaemonism, but rather to the psychology assumed by the challenge to justice in the *Republic* itself, according to which because we want our own good, when we do things contrary to our own good, what we do is not only unwilling but also has its source in some power external to us. Soul-division requires Socrates to take back one thing he has said in his inferences from our natural desire for our good in earlier dialogues, namely, that *we do not desire (epithumein)* those things that we believe are good when in fact they are bad.

On the minority interpretation according to which *Republic* IV maintains the psychological eudaemonism of the *Protagoras* that we do whatever we do believing it to be the best of the things we can do (making akratic action impossible), soul-division identifies good-independent but non-executive desires that can influence the good-directed and executive reasoning part of the soul by causing it to have appearances of, and form beliefs about, good and bad. Like the mainstream view, this interpretation takes it that it belongs uniquely to the reasoning part to be good-directed, but this seems largely based on the assumption that some one thing in the soul must continue to be the source for ‘Socratic’ guise-of-the-good motivations. This is because to account for Socrates’ claim, ‘every soul pursues the good in everything it does, and does everything for its sake’ (*Republic* 505d–e), this interpretation invokes the reasoning part as the executive in every action.<sup>16</sup> Although this solution also excludes non-rational souls from the scope of ‘every soul’, it preserves the intelligibility-conferring advantages of psychological eudaemonism while accommodating the commonsense idea that sometimes our non-rational motivations have an effect on what we humans do.

However, not every case of conflict in the *Republic* can be understood as requiring reason’s endorsement for the action that is done. Most clearly, Leontius looks at the corpses because he has an appetite to do so, and curses himself (or his eyes or his appetite) because he is angry with himself. There is no mention of reason’s involvement in the case, which is introduced as a case of spirit’s conflict with appetite. To get reason involved, something like the following seems to be required:

'we can see at least two distinct moments: (1) one of internal struggle, where the *epithumêtikon* and *thumos* pull the soul in opposite directions *hama* . . . (2) an executive moment, when the agent looks at the corpses, "overpowered by his desire", while expressing anger . . . [R]eason may have stayed neutral during the struggle at (1) . . . [I]n (2), the agent may choose to look at the corpses on the basis of reason deciding at that point that it is overall better to let the interests of desire prevail over those of *thumos*.'<sup>17</sup>

In (2) Leontius performs two simultaneous actions, looking at the corpses and cursing himself or his appetite or perhaps his eyes, and this interpretation requires the reasoning part to be the executive for both; but since the two actions are opposed, involving simultaneous desire for and aversion to looking at corpses, the executive reasoning part would have to be divided. And a division within the reasoning part is a good deal more problematic than one within the appetitive part, since virtue is rule by reason. One might try to restrict 'doing' and 'action' so as to avoid this outcome (for example, one might say that the looking is not an action but the cursing is), but *Republic IV* is quite inclusive about the things the soul 'does': they include learning, growing angry, and desiring (436a).

*Republic X*'s arguments provide conclusive evidence that, contrary to the minority interpretation, in the *Republic* Socrates accepts the possibility of action contrary to one's belief about what is best. Whereas *Republic IV*'s soul-division is based on conflicting desires, *Republic X*'s first soul-division is concerned with conflicting beliefs.

- (C1) The same magnitude appears through sight not to be equal close up and far off; the same sticks look bent in the water and straight outside it . . . (602c).
- (C2) Measuring, counting, and weighing help us so that (for example) the apparent large or small doesn't rule in us<sup>18</sup> but the calculating, measuring, and weighing thing does (602d).
- (C3) Calculating, measuring, and weighing are the work of reasoning (602d).
- (C4) Often after the reasoning part has calculated, the opposites (either (a) to its calculations, or (b) to the apparent large or small<sup>19</sup>) appear to it [*toutôi*, the reasoning part] (602e4–6).
- (C5) The same thing can't believe opposites about the same thing at the same time (602e).



- (a) assuming that the soul forms [contrary] beliefs on the basis of the appearances and the calculations; or
- (b) assuming that after calculating, the reasoning part is the subject of new appearances, rational ones on the basis of which it forms beliefs, but another part continues to believe contrary to these.
- (C6) So the thing that believes contrary to measurement can't be the same as the thing that believes in accordance with measurement (603a).
- (C7) That which believes in accordance with measurement is the best thing in us.
- (C8) That which opposes it is one of the inferior things in us (603a).

Unlike Book IV, Book X is concerned with a bipartition rather than a tripartition of the soul, because its main concern is the conflict between following and opposing calculation. I believe this division folds spirited motivations (which, I'll argue in Chapter 5, always follow the reasoning part) in with rational ones. We find some confirmation of this in the way the next, motivational, argument for soul-division describes the best part, namely, as not only reasoning, but also following reason and law:

- (M1) A decent person who has lost his son will grieve moderately, resisting his pain especially in public (603e–604a).
- (M2) Reason and law bid him to hold out against his pain but the experience itself (*auto to pathos*) drags him to it (604a).
- (M3) When an opposed course of action (*agôgês*) arises in the same person concerning the same thing, at the same time, we must say that he has two [parts] (604b).
- (M4) The law says, don't overreact: this may turn out well or badly in the end; what's to come isn't improved by taking this hard; human affairs are not serious; grieving impedes deliberation about what to do (604b–c).
- (M5) The part that is ready to obey law and calculation is the best (604d).
- (M6) The part that leads us to grieve is irrational, lazy, a friend of cowardice (604d).

If we can form beliefs contrary to reasoning, as (C2) and (C4) indicate, and we can grieve contrary to reasoning, as (M6) indicates, there doesn't seem to be any good reason to think we aren't able to act (for example, look at corpses, drink unhealthy drink) contrary to reasoning. What we have to let go of is the view, foreign to the text, that reason is the executive of every action, or the proper subject of every belief about what to do. It is also worth noting that (as argued in Chapter 2) while the dialectic in the *Protagoras* provided some motivation for Socrates to redescribe akrasia as unstable belief under the influence of appearances, because if true, it would show how virtue could be knowledge and teachable, the dialectic of the *Republic* does not provide such a motivation.

Our discussion so far shows that to *Republic* IV's argument that we have in our soul multiple sources of motivation, which originate in calculation or bodily states, *Republic* X adds that these motivational sources are either calculation-following or calculation-indifferent and appearance-led. (M4) in particular enriches the notion of calculation to include taking a broader view of one's experience, and this corresponds to the Book IV claim that the reasoning part of the wise person has knowledge about what is good for each part of the soul and the soul as a whole (442c). The fact that the reasoning part has the ability to calculate what is good for the whole, or all things considered, needn't suggest that the other parts are good-indifferent, only that their outlook is limited.<sup>20</sup> But to say that the outlook of something is limited is to say it falls short, not to say that it fails to aim at what the comprehensive outlook succeeds in capturing.<sup>21</sup>

My interpretation still needs to (i) explain *how* all the soul-parts are good-directed and (ii) harmonize the possibility of akratic action with the claim that 'every soul pursues the good in everything it does' and 'does everything for the sake of the good' (505d–e). Specifying the sense in which all psychic parts and activities are good-directed requires further consideration of the characterization of the parts in the *Republic* (Section 4.3) and *Timaeus* (Section 4.4).

### 4.3 Characterizing Soul-Parts as Good-Directed: Agent-Likeness and Personification

It needs to be admitted at the outset that the *Republic* is not rich in descriptions of the goodness or good-directedness of the non-rational soul-parts. But this is because the *Republic* is devoted to showing the

preferability of justice, which is reason's rule over the irrational, to injustice or insubordination by the irrational. Appetite and spirit are not good at ruling a person, and since the discussion is concerned to evaluate each soul-part as ruler in a soul, it is inevitable that the lower soul-parts should not come off looking good. But this does not mean that they are not good-directed at all.

The way the *Republic* represents soul-parts' good-directedness is to represent each soul-part as an agent pursuing the good under some conception: knowledge, honour, or bodily pleasure. Because some commentators have found the soul-parts' agent-likeness unfortunate, I need to establish that agent-likeness is an important feature of Plato's account of soul-parts, and to defend it, before turning to its use in representing soul-parts' good-directedness.

Although *Republic* IV introduced the three elements in the soul first by their distinctive functions—learning, growing angry, desiring bodily pleasures—we've seen that the three soul-parts have some functions in common. For example, each soul-part is able to form beliefs, including the moderation-constituting belief that the reasoning part ought to rule (442c–d), the contrary-to-measurement belief that the stick in the water is bent (602e), and the contrary-to-calculation belief that the death of a loved one ought to be lamented (604b–d). Further, each soul-part desires, and takes pleasure in, some characteristic object—the reasoning part, wisdom; the spirited part, honour; the appetitive part, bodily pleasure (580d–581c). Plato seems to model the soul-parts on the citizens of different classes in the city, who differ in their pursuits and competences, rather than their basic capacities. The agent-likeness of soul-parts is the basis of Plato's characterization of virtue as psychic harmony—a condition in which not only does each part do its own work, but also all three parts share the same belief about which part should be in charge—and vice as psychic disharmony, a condition in which the non-rational parts usurp the rightful place of the reasoning part in ruling a person's life.<sup>22</sup>

Plato's characterization of soul-parts as agent-like is intentional, rather than inevitable or an unintended consequence of some other philosophical or artistic decision. While one might imagine that it is natural for us to think and talk about our various motivations as if they issued from separate persons within,<sup>23</sup> Plato often finds ways to talk about our motivations without doing so; for example, Socrates likens the

spirited part of the soul to metal in need of tempering (411b, cf. the depiction of the soul as a puppet pulled by strings made of different metals, *Laws* 644c–645a), and he likens the embodied soul to the barnacle-encrusted and maimed sea-god Glaucus, where the accretions may be the soul-parts or the motivations that draw on them (611b–612a). As we saw in Chapter 3, in the *Gorgias*, he likens the appetitive part of the soul to a jar that may be leaky or sound and the reasoning part either to a sieve or something less holey (493a–494a). One might also imagine that the characterization of soul-parts as agent-like is encouraged by the city–soul analogy, since if justice is reason’s rule, in the soul as in the city, it is easy to dwell on the similarities between people and soul-parts. But other dialogues feature agent-like soul-parts without the city–soul analogy. So, for example, in the *Timaeus* the appetitive part of the soul, although deprived of belief (77b–c), is capable of either obeying reason or refusing to obey it (70a), and of practising divination (71d–e); in the *Phaedrus*, the soul is composed of a charioteer (the reasoning part) and two horses, one horse (the spirited part) being a lover of honour with modesty and self-control, guided by verbal commands and controlled by a sense of shame (253d–e, 254b–c), while the other (the appetitive part) bargains and bullies in order to have sex (254c–d). Furthermore, in the *Phaedo*, where he does not divide the soul but opposes soul and body, Socrates describes soul and body as agent-like, so that, for example, the soul counters and masters bodily affections by means of threats, admonitions, and violence, talking to its desires, angers, and fears as one thing talking to another (94c–e).

It’s a puzzle just how agent-like soul-parts are. Does Plato think that the ability to calculate means to an end is the unique ability of the reasoning part, or does the agent-likeness of all the parts imply that each of the parts can engage in means–ends reasoning, as might be suggested by the talk of parts persuading each other, and of injustice being a matter of non-rational parts taking the place of reason? On the one hand, in *Republic* IX, Socrates calls the appetitive part ‘money-loving’ and says that its pleasure and desire are for profit (580e–581a); some interpreters take this to mean that the appetitive part comes to love money by calculating that money is the means to the enjoyment of bodily pleasures. And when Socrates describes the oligarchic man’s appetitive part enlisting the reasoning part to calculate how best to acquire money (553c–d), these interpreters argue, it does so on the

basis of its own calculation that the reasoning part is the best discoverer of means to acquiring money.<sup>24</sup> On the other hand, in *Republic* IX, Socrates justifies calling the appetitive part money-loving on the grounds that money *is* (rather than is believed by the appetitive part to be) the means to the satisfaction of its appetites (580e–581a). In *Republic* IV Socrates says that it belongs uniquely to the reasoning part of the wise person to be able to calculate what is best overall (442c), and in *Republic* X, that it belongs uniquely to the reasoning part to do so on the basis of measuring and weighing rather than on the basis of appearances (602d).<sup>25</sup> However, perhaps these statements only make some means–ends reasoning that meets fairly high standards unique to the reasoning part, without ruling out from the ability of the lower parts the sort of reasoning that determines that such-and-such is a means by which it can achieve pleasure, or the sort of reasoning that makes such a determination by comparing larger and smaller magnitudes of pleasure.

Yet if each part can calculate, conflicts between what seems best to a single part in the short as opposed to the long term loom, threatening the unity of each part.<sup>26</sup> Against this, if only the reasoning part can determine means to ends, then some instrumental desires seem to end up homeless. In *The Brute Within*, Hendrik Lorenz argues that since only the reasoning part can calculate, the appetitive part cannot value money as an instrument, but may retain, perhaps in memory and imagination, complexes of desire–action representations, so that its attitude towards the bodily pleasures money can buy leaks over into an attitude to money (50–1). But consider the akratic smoker, who, rationally averse to smoking but overcome by the appetitive desire to smoke, must go to the ATM, withdraw cash, go to the gas station across the street, and buy cigarettes. By hypothesis, her reasoning part, being averse to the end of smoking, must be averse to actions performed for the sake of smoking. Is it plausible that her appetitive part's desire–action memories explain her going to the ATM, withdrawing cash, going to the gas station, and buying the cigarettes? Such a person might, if the ATM is broken, use her credit card, or borrow cash from a friend, or scrape together her change, or steal, in the process performing a new instrumental action for the sake of getting the smoke. It seems ad hoc to insist that in such a case any instrumental action is merely remembered whereas in a case where reason owns the ends, instrumental actions are calculated.

Agent-like soul-parts have been criticized for reproducing, at the level of explanans, all the behaviours in the explanandum. Julia Annas puts the problem succinctly: 'The parts are explanatory entities, parts needed to explain the behaviour of the whole. If they themselves, however, can be explained in the way that the whole person is, have we not reproduced the problems that led to the need for the theory in the first place?'<sup>27</sup> But if the parts are simpler than the whole person, then we will have made explanatory progress. Setting aside the vexed issue about instrumental reasoning, it does seem that soul-parts are simpler than the whole person in the following way: whereas we as complex persons value bodily pleasure, the fine or honourable, and knowledge of the truth, we may think of each of the parts as a psychological eudaemonist, in each of its movements impelled towards the good under only one of these conceptions.<sup>28</sup>

What is it that soul-parts are to explain? In *Republic* IV the focus is on actions—why we do the things we do; why we sometimes don't do what we should and seem to know we should—and in *Republic* VIII and IX on characters—both what someone is like and how she comes to have a character of that type. Socrates describes each character-type in terms of a ruling conception of what is good, which the agent pursues in her actions, and which the agent has in virtue of the soul-part ruling in her soul.

It's worth studying how Plato conceives of a soul-part's rule. In the *Republic* Socrates sometimes describes the relationship between soul-parts in terms of force—which part is stronger than which—and sometimes in terms of shared or conflicting outlooks. *Republic* IV already characterizes having an appetite both psychologically, as a pro-attitude, and as a force or impulse.<sup>29</sup> These attitudes and/or motions explain our actions and our conflicts both in the way of the Many and in the way of Socrates in the *Protagoras*. For example, Socrates says the soul of the thirsty person, insofar as he is thirsty, 'wants (*bouletai*) to drink' and 'is impelled (*hormai*) towards' drink; he describes the virtue of self-control both as a condition of control (*enkrateia*, 430e7) over the pleasures and appetites, which means the better part 'is stronger than' (*enkrates êi*, 431a6) the worse, and as a condition in which all the parts 'share the same opinion (*homodoxousin*, 442d1)' that reason should rule. These double characterizations raise the question: are the *Republic's* explanations in terms of soul-parts' judgements of goodness and in terms of forces or motions alternative ways of explaining the same phenomena (as

in the *Protagoras*, where explanation in terms of appearance- or knowledge-based judgements of goodness replaces explanation in terms of forces that push and pull the agent by their strength), or do appearances and judgements explain some actions but forces and motions explain others? That Plato should think the former is supported by the fact that in *Laws X* the Athenian identifies various psychological acts with motions of the soul that originate all the motions in bodies:

[T]he soul moves (*agei*) everything in the heavens and on earth and in the sea by means of its own motions, the names for which are wishing (*boulesthai*), inquiring (*skopeisthai*), caring (*epimeleisthai*), deliberating (*bouleuesthai*), true and false judging (*doxazein*), being pleased and pained (*chairousan lupoumenên*), being bold and fearful (*tharrousan phoboumenên*), hating and loving (*misousan stergousan*). . . . (896e9–897a4)

So it does not seem foreign to Plato to ask: can all talk of pushing, dragging, being impelled, and being stronger than be cashed out in terms of psychological acts that involve representing some object as good or bad, and vice versa, or does Socrates' treatment of phenomena like belief, emotion, and action contrary to calculation in the *Republic* require him to supplement judgement with forces for certain phenomena?<sup>30</sup>

The first of these possibilities, that explanations in terms of forces or motions of varying strengths may be replaced by explanations in terms of judgements of goodness, is different from the minority account of soul-division discussed in Section 4.2, because it does not assume the reasoning part as the executive for every action. Instead, it allows that any part can be the executive. But that there are still two ways in which a soul-part can determine what the agent does is suggested by Socrates' statement that the oligarchic character uses force (*biai*, 554d1) to get his appetites to comply with his better desires by compulsion and fear (*anangkêi kai phobôi*) rather than by persuading them that it's better not to act on them (554c–d). We might call the persuasion kind of relationship between the executive part and the others 'hegemony'. In this context, the other relationship, of compulsion and force, does not describe just any pushing and pulling on the force model but a defective way of commanding and obeying. It's important that a psychological account in terms of judgements recognizes these two ways of executing a judgement, for the main case for soul-division was that it was needed to account for simultaneous

opposed attitudes towards the same thing, be they desires and aversions or beliefs and contrary beliefs.

The second possibility, that the motivations' varying strengths supplement evaluative judgements to explain only certain beliefs, emotions, and actions, may seem ad hoc. When do judgements explain and when do strengths? If we can't do without strengths in cases of conflict why are we so sure strengths aren't the explanation in the other cases, with judgements of goodness only rationalizing, or justifying, psychological outcomes that are caused by the strength of the agent's various motivations?

A closer examination of individual soul-parts as psychological eudaemonists suggests a reason why explaining some actions as due to judgement and others as due to the strengths of the forces in the soul may not be ad hoc. In the *Protagoras*, psychological eudaemonism explains all choice of the worse option as due to ignorance, remediable by acquisition of an art of measurement, because goodness is scalar—as Socrates puts it there, the distant pleasure and the proximate one don't differ in any way other than by pleasure and pain (356a–b)—and correct measurement of pleasure guarantees correct identification of the greater quantity of pleasure. But ignorance does not keep the agent from trying to get the greater quantity of pleasure, and that is how a mistaken appearance that X is the greater pleasure can lead the agent to choose X: according to psychological hedonism, the agent, whether ignorant or knowledgeable, isn't aiming at anything other than the greater pleasure. But now, if we introduce another good, say wisdom, it seems that an agent that aims at the good could pursue only pleasure, or only wisdom, or both in some combination, depending on what she conceives the good to be. So if one agent-like part, believing that the good is pleasure, pursues the greater pleasure, and another, believing that the good is wisdom, pursues greater wisdom, and both are members of the same larger entity (the soul whose motions move the body), then, in cases of conflicting evaluation, we may well need to appeal to the respective strengths of the parts to explain which of the two, wisdom or pleasure, the person goes for. In *Republic X*, Socrates speaks of soul-parts being strengthened and ruling as a result of being nourished and watered, or being withered and ruled (606b, d). It seems that a judgement or a motion from a stronger soul-part would itself exert more force on the person than a judgement or motion from a weaker soul-part. This could



explain why the same calculation-based judgement trumps the appearance-based judgement in the case of one person but not in the case of another.

On the interpretation I'm proposing, because soul-parts have different ends, what one part determines is good (because, e.g., pleasant) doesn't automatically appear good (e.g. fine or all things considered good) to the other parts unless there is some mode of translation or common currency between them. Since the reasoning part is able to take the interests of the lower parts into account, and does so if it has wisdom, wisdom presumably includes the development of such a common currency in terms of which it can assess the claims of the different parts. But it is a further question whether and how the reasoning part's judgements might be translated 'down' from the reasoning part into the terms or currency of appetite.

Jon Moline describes wisdom's communication or persuasion of the lower soul-parts in terms of reason having to 'transplant into alien parts its own opinions, or, more accurately, opinions corresponding in content to its knowledge' (15). Reason does this by means of speeches that are charms, which one learns by the study of rhetoric, and which are exemplified in Plato's dialogues by his myths. Stories that soothe and quiet the lower parts (reassuring it that the soul is immortal, for example, and that death is no evil for a philosopher) increase the relative strength of the reasoning part. It's crucial to this account that the non-rational soul-parts must see the benefits of virtue in their own terms: pleasures, in the case of the appetitive part, and honours, in the case of the spirited part. On this way of conceiving soul-parts' relation to their objects of desire and pleasure, the psychic harmony that is justice would have to be the result of reason's superior ability to persuade the other parts—not that the good is something other than what they characteristically desire, but rather about which things in the world do or don't give them the maximum bodily pleasure or honour that they seek. And the extensive pre-rational education described in *Republic* II–III would be designed to make the non-rational parts receptive to these judgements of the reasoning part.

Yet it seems that education and corruption as envisioned in the *Republic* require greater flexibility in the objects of desire and pleasure of the soul-parts than this. For if a soul-part desires all and only bodily pleasures, how can it be in a harmonious condition when it only gets

reason-approved pleasures (586d–e)? Won't it still be, very unharmoniously, craving all the other pleasures? And conversely, how is it plausible to say that even in the soul of the tyrant, the reasoning part seeks wisdom? In what does the tyrant's extreme of corruption consist, if not in that even his reasoning part believes that some lawless desire ought to be fulfilled? Further, how could Socrates call the appetitive part money-loving unless that part could acquire the desire for money (whether as a result of calculation or association), since we certainly aren't born with the desire for money?

An alternative to Moline's picture suggests that bodily pleasure, honour, and wisdom, identified by Socrates as the characteristic (*idia*, 580d8) pleasures and objects of desire of the three parts, are not what soul-parts desire by nature. Instead, bodily pleasure, honour, or wisdom are what a person dominated by appetite, or spirit, or wisdom, would pursue—so that they would be distinctive but not inevitable pleasures and objects of desire of the three parts, with which pleasures, honours and learning each part seeks being determined by the person's character.<sup>31</sup>

In favour of this alternative, it seems that the relationship between parts and functions is not only one of overlap, but also depends on the agent's character, or state of soul. In the *Phaedrus* the procedure for giving an account of the soul is to say of each simple what are its powers (*dunameis*) to act and be acted upon, and on what and by what. In *Republic* V, Socrates defines a power (*dunamis*) by what it does (*ho epargazetai*) and what it is set over (*eph' hōi esti*) (477c). For example, knowledge cognizes infallibly and is set over what-is; opinion cognizes fallibly and is set over what-is-and-is-not. Socrates takes some pains to emphasize that functions and objects are unique to a power. It would seem that their respective objects give knowledge and opinion their characteristics: knowledge is stable because of the perspectivelessness and permanence of what-is, and opinion is unstable because of the perspectival and impermanent nature of what-is-and-is-not. But both knowledge and opinion are powers of the reasoning part of the soul in different conditions. Other examples of powers in the *Republic* are justice and injustice (351e, 358b, 367a, 443b, 588b), civic courage (429b–430b), dialectic (521d, 532c, 533a), crafts (346a–b), and sight (508a–509b). At least the first two of these are powers of the soul as a whole, requiring that each part be in a certain condition.<sup>32</sup> What consideration of the soul's powers tells us is that we must be careful,

in characterizing the parts, not to ascribe to a part without qualification a power or function or characteristic that belongs to that part in only some but not other conditions.

In favour of Moline's picture, Socrates says that it is by force that the oligarch gets the reasoning part only to seek the truth about how to make profit and the spirited part to have ambition only for wealth: he speaks of the appetitive part causing the reasoning and spirited part to be seated on the ground beneath it and enslaved (*parakathisas hup' ekeinô kai kata-doulôsamenos*) (553d). If a slave is someone who must be compelled to do the bidding of her master, this suggests that rather than the oligarch's reason and spirit having new or restricted objects of desire and pleasure because they have been corrupted, they continue to desire the truth generally and all sorts of honours, but are not permitted to pursue these. And if corruption could achieve appetite's hegemony over reason and spirit, with the latter's acquiescence contributing to the legitimacy of its rule, there would be no need to use force.

But perhaps the oligarch must use force on his reason and spirit and is pulled in several ways only because he is midway between the virtuous and the most vicious. Socrates says that in the democratic soul (which is worse than the oligarchic), bad desires occupy the empty 'acropolis' (560b), and that the democratic man gives over rule (*archên*) to any desire that comes along, not admitting any truth into his guardhouse (*to phrourion*) (561b). If this is to say that his reasoning part has been hijacked by the appetites, and, having ceased to desire wisdom, now desires appetitive gratification, then it would seem that in the democrat, being ruled by appetite is not just the appetitive part being the strongest or biggest force in the soul, but also the hegemony of appetite: the acceptance by the other parts of its desires as their desires. In the oligarchic soul, by contrast, appetite would seem to rule by force rather than hegemonically.<sup>33</sup>

#### 4.3.1 Personification

In addition to being characterized as agent-like subjects of appearances/beliefs and desires that make them self-sufficient to produce actions, each of the soul-parts in Plato has a persona; that is to say, a human or animal character.<sup>34</sup> This is vividly illustrated in *Republic IX*'s image of the human soul as composed of a human being (the reasoning part), a lion (the spirited part), and a multi-coloured beast with heads of gentle and

savage animals (the appetitive part). The person who wants to be just must have the human being in control, feeding and domesticating the gentle heads and preventing the savage ones from growing, making the lion its ally and bringing about friendship between all three kinds (588b–589b).

As we have seen, Plato sometimes treats the soul's independent motivations without representing them as agent-like, so it's fair to ask, why does he, in addition to treating them as agent-like, personify them? Whereas agent-likeness explained the parts' impulses, and the actions they motivate, I suggest that personification solves a practical problem about how to respond to the condition of having conflicting beliefs or conflicting desires. Of course the best way is to examine the contents of our beliefs and desires and determine which of any two conflicting beliefs or desires is correct, but when one is not in a position to do this, a rough-and-ready way to respond is to consider the belief's or desire's likely source in a part, because parts have ends and this makes them especially sensitive to certain considerations but perhaps not others. Personifying soul-parts enables an evaluative classification of psychological attitudes without too-detailed an involvement in their contents, and such evaluation-at-a-distance seems especially useful for the less-than-fully-wise.

For example, suppose I feel angry at the university administration. I can either assess the appropriateness of my anger, in light of the administration's indifference to substantive educational goals and kowtowing to the businesspeople on the Board of Trustees (Is my level of anger justified? Does deference to the businesspeople serve some good end? Will it do more good than harm to act on my anger?), or I can say to myself, 'There goes the always-angry part of me again', thereby putting some distance between myself and the anger. The reason to do the latter would be if in general (not in this specific case, where I'm supposing direct assessment is not possible) anger is of no benefit to me or anyone else, so there is no point to the more-involved assessment. Constructing personae on which to hang my attitudes gives me another way of managing them, and perhaps of intensifying my identification with or alienation from them. Of course, this is a rough-and-ready form of management, since my anger might be appropriate in the circumstances, but my longer-term goal, of rational control, may be better served by the rough-and-ready approach now.<sup>35</sup>

One way it's easy to personify soul-parts is by the goods that appear to them, because this puts one in mind of a type of life or a type of person or animal. We imagine people with a narrower or a more enlightened understanding pursuing truth, honour, or bodily satisfactions as good, and the argument among the wisdom-lover, honour-lover, and profit-lover as to whose life is most pleasant, which follows *Republic IX's* identification of soul-parts by characteristic pleasures (581c–582e), encourages us to do just this. Our sense of soul-parts as each pursuing their objects *as good* is the result of such personification.

Plato's characterization of the lower soul-parts has been criticized on the grounds that they don't really seem to be part of the self, with the result that the person isn't responsible for having their attitudes, only for how she rationally deals with them.<sup>36</sup> My answer to this is two-fold. On the one hand, it seems that by locating the origins of even our bad actions in our soul, and saying they are our desires, Plato is making the subject of such attitudes responsible for them more than do his predecessors (as we saw in Chapter 3). On the other hand (as we also saw in Chapter 3), it is true that the primary way in which we are responsible for our attitudes is forward-looking: given that we have them, what can we do to discipline them so that our actions achieve the good? But this is not obviously mistaken. It is worth noting the wealth of evidence collected by contemporary cognitive psychologists that our minds contain independently functioning modules that are somewhat sealed off from our reasoning processes, or 'informationally encapsulated', so that what we have to do with the attitudes they generate is to manage their bearing on our actions rather than to change them. For example, even if we know that the conjunction of two events is less likely than either of those events by itself, it can still feel to us that 31-year-old Linda, outspoken and bright, who studied philosophy and participated in anti-nuclear demonstrations, is more likely to be a bank-teller and a feminist rather than just a bank-teller.<sup>37</sup> This appearance, like the contrary-to-calculation visual appearances Plato discusses in *Republic X*, is something we are responsible for only in the sense that we can manage it so that it doesn't inform our actions. Of course, Plato includes desires among the attitudes that we can rationally manage and one might think we are responsible for having and not only managing those desires, but presumably this isn't true of desires that originate in bodily conditions or in our upbringing under faulty constitutions.

## 4.4 Teleological Explanation and the Divided Soul

In Section 4.3.1, I argued that the *Republic's* three soul-parts are consistent with the claim that 'every soul pursues the good and does everything for its sake' (*Republic* 505d–e) because each soul-part pursues the good under a more-or-less adequate conception (pleasure, honour, overall goodness). In the terms introduced during our discussion of the *Protagoras* in Chapter 2, each soul-part is a psychological eudaemonist. But, one might object, surely Plato can't think it's literally the case that lower soul-parts represent their objects of pursuit as pleasant or honourable (as opposed to our conceiving of them as such from the outside). Does the (appetitive) soul of a plant represent water and sunlight to itself as pleasant?

The *Timaeus* gives a different, but I think compatible, account of how all kinds of souls or soul-parts might be good-directed. Here, Timaeus provides an account of how our cosmos, we ourselves, and especially our souls have come to be as we are by showing how each thing is for the best. The Demiurge, intelligent craftsman of the cosmos, free of malice and so out to make a cosmos as perfect as possible, makes the world soul by blending Sameness, Being, and Difference so as to enable it to judge and know (37a–b, 41d, cf. *Theaetetus* 186e); thereafter, he makes our rational souls out of the same ingredients (41d) and, before we are embodied, tells us the laws of destiny: rational souls are to be implanted into bodies, as a result of which they will come to have sense-perception, then desire along with pleasure and pain, fear and anger and whatever goes along with these, and their opposites. A soul that overcomes these passions by reason is just, and eventually returns to its assigned star and lives a life of happiness in accord with its character; a soul that doesn't is unjust and is reborn into an animal resembling its character again and again until it gains rational control of its passions (41e–42d). The effect of sense-perception, desire, and the passions on the rational soul is to block its capacity for knowledge and produce in it false judgements (43d). The Demiurge leaves it to the lesser gods to execute mortal embodiment (41b–d, 42e–43a), and what they do is to design bodies which house souls' distinct activities in appropriate bodily locations (69c–71e). In the case of human beings, the gods separate the place of the rational soul, a round and thinly covered head atop the body, from

the places of the spirited and appetitive soul-parts, by a neck; they station the spirited emotions at the heart so as to facilitate the transmission of its threats and exhortations through the blood vessels to all the limbs; they locate the appetitive desires in the abdomen, near the organs of digestion and reproduction, and fashion a liver that can represent thought in images that stir up pleasure and pain. All this is so that the rational part of the soul can think in peace, without the disturbances of appetite and emotion, and thereby regain control. Animal bodies too are designed to accommodate the various kinds of psychic lives appropriate to psychic characters; for instance, the many feet of land animals orient their heads downwards since they no longer make use of their rational motions (91d–92c).<sup>38</sup> The gods' design of mortal bodies and mortal soul-parts is, like the Timaeon cosmos as a whole, 'psychocentric': everything is arranged for the good of the soul.<sup>39</sup>

If the central goal of the lesser gods' design of bodies and soul-parts is that the rational soul be freed to engage in its characteristic activity—reasoning about the truth—in peace, they have good reason to make us able to act from appetite or emotion without the 'online' involvement of the rational part, and so without its judgements of good and bad.<sup>40</sup> Just as it would impede the goal if each rational soul had to reason out all its bodily mechanisms, the goal would seem to be impeded by reason's involvement in each psychic act of emotion and appetite. Of course, in addition to being optimal, it has to be possible for acts of emotion and appetite to take place without reason's involvement, but we see the possibility of independent activity by the lesser soul-parts already in the plant that grows in the direction of sunlight, equipped with just perception and desire (77b–c); why would it not be possible for an adult to eat because of hunger without also thinking, 'The best thing I can do now is to eat'? Timaeus says that reason needs to be in control, but on my proposal, reason's control involves some delegation of responsibility for actions to well-trained appetites and emotions so that it doesn't have to busy itself with each action.

The *Timaeus* indicates that bodily experiences are a powerful source of false beliefs about the value of appetitive goods. Still, it must be possible to have those experiences without succumbing to the false beliefs, since it is possible to achieve rational control while embodied.<sup>41</sup> Our having the ability to put our responses to hunger and thirst and such things on 'automatic pilot' would seem to facilitate living without making false judgements on the basis of bodily experiences.

If this proposal about our psychic design is right, it would seem to follow that our false judgements and infelicitous desires belong to us insofar as they are a consequence of our possession of soul-parts dedicated to our nourishment and protective responses. And in this case, it won't be correct to say we 'don't believe' falsehoods or 'don't want' bad things, even on the occasions on which they are 'unwilling' because contrary to our real good. For they are not contrary to our nature, the various parts of which are designed to go after objects of specific kinds, serving our real, overall, good when regulated by reason.

On this sketch of the psychology of the *Timaeus*, 'every soul pursues the good and does everything for its sake' (*Republic* 505d–e) in the sense that every soul is set up by the gods so that it can perform its characteristic functions. In the human case, this includes the careful exercise of reason in judgement and inquiry. Reason, because it can judge or form beliefs about how things are, is better fitted for attainment of our good than is appetite, but even appetite, responsive to sense-perception and pleasure and pain (77b–c; and presupposing some cognitive contact with what fulfills it (*Philebus* 35a–c)) is set up to achieve some good: that is how the appetitive soul of plants, inclining them to light and nourishment, serves their good. But just as every soul's pursuit of the good is compatible with different kinds of souls having different means of pursuing the good, so too, it seems possible that every action's being done for the sake of the good is compatible with different actions being motivated differently than by the agent believing that it is good or representing it to herself as good. Even soul-parts that pursue and act for the sake of achieving bodily pleasure, we can now see, are pursuing and acting for the sake of some good—not only because pleasure is a good, but also because it is good that there be plants, and good that the souls of animals are specialized so that in addition to nourishing themselves and growing, they can pursue higher goods.

## Notes

1. How should we distinguish 'in the same respect' (*kata t'auton*) from 'in relation to the same thing' (*pros t'auton*)? To avoid redundancy, Adam renders *kata t'auton* by 'in the same part', citing *Sophist* 230b as a parallel (J. R. Adam (ed. and comm.), *The Republic of Plato* (Cambridge, 1902), vol. 1 at 246–7). That text says: when refuting someone, the refuters set his opinions



side by side, and show they are opposed to each other at the same time (*hama*) on the same subjects (*peri tôn autôn*) in relation to the same things (*pros t' auta*) and *kata t' auta*. But this isn't decisive between 'part' and 'respect', for Socrates could be reserving *pros t'auton* for relations to something external to the thing being considered, as in *Theaetetus* 154c–155c, where the six dice are greater *pros* the four and lesser *pros* the twelve. R. Stalley argues that since the axis and circumference aren't parts of a spinning top, Plato must mean 'in respect of' by *kata* ('Plato's Argument for the Division of the Soul', *Phronesis*, 20 (1975), 110–28); it would be open to him to accept the distinction I've proposed between *kata* and *pros*. In our *Republic* passage, Socrates adduces two cases to illustrate the principle of opposites. In the first case, having affirmed that it is impossible for the same thing to be moving *hama* and *kata to auto* (436c), Socrates explains that if a person is standing still and moving his head and hands we should say *to men ti autou estêke* (something of him is standing still), *to de kineitai* (the other is moving). This is a case where the person turns out to be *many* (by having parts rather than, for example, being a succession of entirely distinct persons, as 'Protagoras' proposes at *Theaetetus* 166b in order to save 'knowledge is perception' from 'the same person knows and doesn't know' given cases in which a person with closed eyes remembers what he doesn't see (163d–164d) or a person with one eye closed sees and doesn't see (165b–c)). In the second case, a spinning top is said to stand still *kata to euthu* and move *kata to peripheres* (436d–e). The spinning top isn't a case of 'many' (its resolution doesn't involve a *to ti men autou/to de*). At *Metaphysics* IV.1005b 20, Aristotle states the principle of opposites by specifying 'simultaneous' and then treating the qualification '*kata to auto*' as a catch-all: 'however many distinctions we should make in relation to the logical difficulty, let them be made'.

2. Because Plato is so metaphysically non-committal, interpreters disagree about the kind of thing a soul-part is. Two important interpretations very different from my, in this respect completely mainstream, approach are C. Shields, 'Plato's Divided Soul', in M. McPherran (ed.), *A Critical Guide to Plato's Republic* (Cambridge, 2010), 147–70, who argues that we should think of the parts as properties that souls can have but need not: e.g. 'being appetitively desiring' is an aspectual part of a soul that can disappear; and J. Whiting, 'Psychic Contingency in the *Republic*', in R. Barney, T. Brennan, and C. Brittain (eds), *Plato and the Divided Soul* (Cambridge, 2012), 174–208, who argues that it is defective souls that are partite.
3. Cf. Antiphon, *Truth* 7b, discussed in Section 3.3.2.
4. As noted by Myles Burnyeat in 'Culture and Society in Plato's *Republic*', *The Tanner Lectures on Human Values* (Salt Lake City, 1999).

5. Grube-Reeve (in Cooper and Hutchinson, *Complete Works*) translates 'and', suggesting that the point is that insofar as one has an appetite for something, all these impulses follow; perhaps, though, an appetite gives rise to only one or more of these impulses, depending on the circumstances (e.g. if one can have what one wants because someone else provides it, one would accept it, rather than pursuing it).
6. Elsewhere, Socrates takes the relativity-in-experience of large and small to raise the definitional question 'what is the large-itself'?—cf. *Phaedo* 96e–101d; *Republic* 523e–524a—whereas Aristotle, in his *Categories* 7 discussion of relatives (things that are what they are in relation to other things, 6a36–37, cf. 8a31–32), places large and small squarely in the category of correlatives for which knowledge of either of the pair exhausts knowledge of the other, and distinguishes such correlatives from relatives in which one of the pair has cognitive priority: contrast 'large', where all it is to be large is to be larger than something smaller, with 'head', where although a head must be of a body, that's not all there is to being a head. For discussion of Aristotle, see D. Sedley, 'Aristotelian Relativities', in J. Brunschwig, M. Canto-Sperber, and P. Pellegrin (eds), *Le Style de la Pensée: Recueils de textes en hommage à Jacques Brunschwig* (Paris, 2002), 324–52. Aristotle's discussion helps us to see that as an *appetite* for drink (rather than just being 'for drink'), thirst is a little different from the correlatives (larger, double) with which Plato lumps it, and is more like a state, the knowledge of which is not exhausted by knowledge of its object.
7. See, e.g., C. Kahn, 'Plato's Theory of Desire', *Review of Metaphysics*, 41 (1987), 77–103 at 85: 'In order to establish the distinction between reason and appetite Plato must here define, for the first time, the notion of a desire that is essentially independent of any judgment concerning what is good, beneficial, or advantageous'; Irwin, *Plato's Ethics*, paraphrases Socrates: 'if [thirst] were desire for drink qua F, it would be desire for F, not desire for drink. It follows that thirst is not desire for drink qua good' (206). Irwin objects to Plato: 'facts about thirst as such do not show that any of our actual desires is a desire for drink as opposed to drink qua good' (207).
8. Carone, 'Akrasia in the *Protagoras* and *Republic*', says: 'Plato . . . is denying that it [viz. 'the Socratic dictum' that everyone pursues the good and on those grounds does what she believes to be best in the circumstances, 117] can be invoked as grounds [cf. *gar*, 438a3] for delimiting the proper object of appetite as such' (118). 'Supposing . . . that all faculties motivating an agent to act tend to their proper objects as good, conflict between thirst and hunger . . . is not . . . explained by these two faculties wanting their respective objects *as good*, but by thirst going for *drink* as good as opposed to hunger going for *food*

as good' (119). 'It is perfectly consistent to claim that thirst qua thirst is for drink while every time we wish to drink we desire drink as good' (120). Reason is distinct because 'only reason can know [or preserve true beliefs about, 130] what is actually good for the overall well-being of the person, and can thus oppose the lower parts of the soul, which have at most narrow-minded beliefs about what simply "appears" to be good to each of them' (120).

9. We might ask why, if the presence of heat can qualify thirst to produce the desire for cold drink, good can't qualify thirst to produce the desire for good drink. I think it can, but not automatically. Rather, qualifying our appetites by our desire for what's good requires that we ourselves do some psychological work to integrate two motivational sources that are somewhat insulated from one another.
10. Possibly the *Phaedo* answers this by attributing conflicting beliefs to the body, which isn't what we are but is our, or our soul's, instrument, and which can somehow encroach on and corrupt us: the body makes pronouncements about what the truth is that can form the soul's beliefs (83d), thereby deceiving the soul (65b) and troubling it and not allowing it to acquire truth and wisdom (66a); the body and its desires cause war (66c). The soul reasons best without the senses (65c); we are freed from the body's folly by dissociating from it (67a); nature orders the soul to rule and be master of the body (80a), by physical means as well as by means of threats and admonishments (94c–e). On this picture, I am my soul, and every pro-attitude towards something bad for/contrary to my soul's natural desire (for truth) is attributable to my body.
11. The appetites include some that are unnecessary, in the sense that we can avoid having them and their satisfaction does not benefit us (558d–559c), and some that are not only unnecessary but also contrary to law, perhaps the law that aims at the good of the whole (571b–d). Natural teleology doesn't show these to be ours, but Socrates' argument blocks us from denying that they are ours.
12. Following H. Lorenz's 'simple' account of partitioning (simultaneous desire and aversion to the same object) in *The Brute Within: Appetitive Desire in Plato and Aristotle* (Oxford, 2006) [= *Brute*], 41–52.
13. Irwin, *Plato's Ethics*, argues that 'unwilling' in (D3) refers to specifically rational aversion rather than aversion in general, on the grounds that otherwise Socrates is assuming without warrant that unwillingness results from reasoning and that all contrariety involves the rational part, an assumption which in any case would be cancelled by the conflict he will want to show between the spirited and appetitive parts (208). However on my reconstruction in (D3) Socrates is stipulating, for this case, or this sort of case, that the unwillingness belongs to the part that has reasoned.

14. To avoid further subdivisions of soul-parts, Irwin proposes that the opposition between appetite and reason is that reason is opposed not just to the object of appetitive desire, but to acting on appetite as such (207–8). I agree that reason can take such an attitude, but I worry that if this is reason's basic attitude, then the teleological account of our having a separate appetitive part would be undermined.
15. Irwin, *Plato's Ethics*, 208.
16. Carone, 'Akrasia in the *Protagoras* and *Republic*', 133. On the alternative that sees some soul-parts as good-independent, 'do everything' (*panta prattein*) need mean only 'exert great effort'.
17. Carone, 'Akrasia in the *Protagoras* and *Republic*', 137–8.
18. By 'rule in us' I suppose Plato means 'determine our choices', rather than merely 'determine our beliefs', since (C4) says that even after calculating the soul can have a calculation-opposed belief.
19. Alternative (b) on 602e1–3 follows Lorenz, *Brute*, 68, which follows Adam's Appendix II to Book X, according to which what reason's appearances are opposed to is given at 602d: the apparently greater or lesser. This is a remote referent, but the connecting thought would be: because of calculation and measurement, a new appearance, reason's, becomes a candidate for ruling in the soul. Alternative (a) asserts the existence/persistence of a contrary-to-reasoning appearance, which on (b) is just assumed, but on (a), the conflicting appearances appear to reason (the referent of *toutôî* at 602e2, in (C4)). Presumably being appeared to isn't one of the attitudes that has an opposite, and conflict in the soul appears only with the formation of beliefs. Still, there is an awkwardness on (a), because one would have expected the appearances to appear to something common in the soul and the two parts to believe in accordance with one of the two appearances.
20. In this I am in agreement with J. Moss, who argues that: (1) this is the most straightforward reading of 505d–e; (2) the definition of thirst as for drink is consistent with the view that being thirsty involves taking drink to be good; (3) the characteristic pleasures and desires of each soul-part are what each part finds worthy of pursuit, to make life worth living; (4) apparent value is just an inferior copy of true value ('Appearances and Calculations: Plato's Division of the Soul', *Oxford Studies in Ancient Philosophy*, 34 (2008) [= 'Appearances and Calculations'], 36–68 at 61–4). However, Moss says that the burden of proof that all soul-parts are good-directed does not fall on the defender of this view, whereas I think we need to do more than show that it is consistent with soul-division for all soul-parts to desire their objects as good; I think we also have to provide evidence internal to the *Republic* that they do desire their objects as good.

21. T. Ganson, 'The Rational/Non-Rational Distinction in Plato's *Republic*', *Oxford Studies in Ancient Philosophy*, 36 (2009), 179–97, argues that all three soul-parts are epistemic agents, with the non-rational parts being non-alethic epistemic agents.
22. J. Moline, 'Plato on the Complexity of the Psyche', *Archiv für Geschichte der Philosophie*, 60 (1978) [= 'Complexity'], 1–26. Moline is followed in this conception of soul-parts by Reeve, *Philosopher-Kings* and C. Bobonich, *Plato's Utopia Recast: His Later Ethics and Politics* (Oxford, 2002) [= *Utopia Recast*].
23. A. W. Price, *Mental Conflict* (London, 1995), 56, speculates about such 'treating parts as persons': 'we may surmise that such conceptions can faithfully capture an aspect of the way the mind pictures itself, a self-dramatizing mode in which it experiences, and transmutes, its own workings.'
24. Moline, 'Complexity'; Bobonich, *Utopia Recast*.
25. For arguments that means–ends reasoning belongs exclusively to the reasoning part, see Lorenz, *Brute*. Moss, 'Appearances and Calculations', adds arguments that perceptions and uncritical evaluative appearances belong to the non-rational parts of the soul. To say that the non-rational parts are uncritical does not mean they don't make judgements—we have just seen Plato use their judgement-making in order to argue for soul-partition in *Republic X*.
26. For discussion, see Bobonich, *Utopia Recast*, 248–52.
27. Annas, *Plato's Republic*, 142.
28. Bobonich argues that the characterization of soul-parts as having both short-term and long-term desires for their own good makes it difficult to avoid further partitioning the soul-parts themselves (*Utopia Recast*, 250–2). For such conflict to arise it would seem that the same soul-part must simultaneously believe that the long-term and the short-term good are both greater, and that requires the same soul-part to form its belief on the basis of both appearances arising from calculation and some other source.
29. Bobonich, *Utopia Recast*, 236–41, distinguishes two ways in which Plato describes desires in this argument: in terms of forces with varying strengths, and in terms of command-and-consent, where the focus is on the content of the desires. Cf. also C. Bobonich, 'Akrasia and Agency in Plato's *Laws* and *Republic*', *Archiv für Geschichte der Philosophie*, 76 (1994), 3–36. Bobonich takes these two not as alternative descriptions of the same phenomena but rather as describing different phenomena.
30. In *Utopia Recast*, Bobonich assumes that according to Plato, the strongest desire determines the agent's action (240, although n. 40 acknowledges that Plato never says this), and that the agent's desires and evaluations may be misaligned. According to Bobonich, Plato thinks that when an agent

akratically does action Y, the agent has (1) a belief that some other action X is best overall, (2) a desire to do X, and (3) a desire to do Y; (3) can determine what the agent does just in virtue of being stronger than (2) (241). It's not clear why there isn't a belief corresponding to (3). It can't be because only the reasoning part has beliefs about what's best (for itself, if not 'overall'), if this analysis is to apply to the case of the spirited part's conflict with the appetitive part—and in any case Bobonich recognizes that a lower part of the soul can have beliefs about what is best for it. Bobonich considers, and rejects, the idea that the lower parts are such that each one's desire for what is best itself is always its strongest desire, on the grounds that this is inconsistent with the unruliness of the lower parts (251). But the unruliness Plato describes is in relation to *reason's* rule.

31. What follows draws on my 'The Powers of Plato's Tripartite Psychology', *Proceedings of the Boston Area Colloquium in Ancient Philosophy*, 24 (2008), 127–50.
32. Complicating matters still further, even powers have some overlap in function. For example, even though Socrates argues that since knowledge is a power distinct from opinion it must have a different object, what-is, rather than opinion's what-is-and-is-not (477c–478b), he also says that guardians and philosophers *know* in what way just and fine things are good (506a, 520c), which implies that there are ways in which these things are not good, so philosophers must know something that is-and-is-not. And he says of himself that he only has *opinions* about the Form of the Good (506b–e, 509c), which would seem to be a quintessential object of knowledge. So it seems that a power is defined by what it does and the object it is set over insofar as the power is *by itself sufficient* to have the object as its object: e.g. knowledge is by itself sufficient to grasp what-is—it does not need perception—and sight is by itself sufficient to grasp colour. This does not mean that no other powers can grasp what-is: opinion plus dianoetic reasoning can grasp what-is, albeit imperfectly. And it's possible for us to know sensibles or opinables by combining knowledge with perception. (Even in the case of sight, although we can't grasp colour by any combination of other powers, we do grasp by sight shape, which we also grasp by touch.)
33. Carone says that what it is for a soul-part to rule is 'to make the whole soul subject to its desires and beliefs' ('Akrasia in the *Protagoras* and *Republic*', 135). Wilburn, 'Akrasia and the Rule of Appetite', argues that psychic rule (rather than *akrasia*) is what is important in the *Republic*, and that a part rules when reason adopts its beliefs. It's worth bearing in mind that both seem to hold that what it is for a psychic part to rule is no different than for a psychic part to triumph in determining a given action—except that the former is more settled.

34. The paragraphs that follow draw on my 'Speaking with the Same Voice as Reason: Personification in Plato's Psychology', *Oxford Studies in Ancient Philosophy*, 31 (2006) [= 'Personification in Plato'], 167–202. However, in that paper I distinguished between partition and personification, lumping together into the latter the two characterizations I distinguish here as agent-likeness and personification.
35. In 'Personification in Plato' I argued for a correlation between personification and facilitating the cultivation of philosophical (as opposed to civic) virtue, thinking that personification would be particularly useful for self-cultivation, but personification would also seem to be useful for evaluatively labelling and managing the attitudes of others. Further, personification isn't necessary for putting distance between oneself and one's motivations; for example, the *Laws* describes the soul as a puppet pulled by strings and exhorts us to make our strings pull along with the golden cord of calculation (644c–45a).
36. J. Annas, *Platonic Ethics Old and New* (Ithaca, 1999), 135–6; Bobonich, *Utopia Recast*, 255–6.
37. A. Tversky and D. Kahneman, 'Judgments of and by Representativeness', in D. Kahneman, P. Slovic, and A. Tversky (eds), *Judgments Under Uncertainty: Heuristics and Biases* (Cambridge, 1982), 84–98.
38. In 'The Soul's (After-)life', *Ancient Philosophy*, 36 (2016), 115–32, I argue that whereas in earlier dialogues souls acquire mortal bodies as a result of their desires, beliefs, or choices, but receive corrective divine punishment or reward while disembodied, in the *Timaeus* and *Laws* mortal embodiment does double duty, both expressing the soul's character and somehow serving to improve it.
39. The term 'psychocentric' is D. Sedley's, from *Creationism and its Critics in Antiquity* (Berkeley, 2007), 126.
40. I argue this case in 'Psychology and the Inculcation of Virtue in Plato's *Laws*', in C. Bobonich (ed.), *Plato's Laws: A Critical Guide* (Cambridge, 2013), 127–48.
41. This needn't be seen as a change of attitude towards embodiment from the *Phaedo*: the *Phaedo* is concerned with the soul's pursuit of truth, a goal for which bodily information is highly inadequate; the *Timaeus*, by contrast, is concerned with the body's instrumentality for a whole range of psychic activities. The difference is that in the *Timaeus* sense-perception also serves a theoretical purpose: sight enables us to see the rational revolutions of the heavenly bodies; sound enables us to hear speech and musical harmonies, all of which become material for the understanding (46e–47e).

# 5

## Why is the Divided Soul *Tripartite*?

### 5.1 Introduction

Chapter 4 suggested a background to explain the *Republic's* interest in arguing that there are multiple independent principles of motivation in human nature, and we saw that the *Republic* identifies the sources of rational and appetitive motivation as calculation or reasoning of some kind, and bodily conditions. But why are there three parts, and why is the third part spirited? In recent decades scholarly understanding of the spirited part of the soul has taken great strides in moving away from the suspicion voiced in the 1970s by Bernard Williams that the spirited part of the soul corresponds to no natural psychological kind but is dictated by the city-soul analogy to be the counterpart to the auxiliary class in the ideal city (thereby allowing the account of civic justice to suggest an account of individual justice).<sup>1</sup> At present there seem to be two main alternative accounts of the spirited part: that it picks out a distinctive set of attitudes that reflect our nature as social creatures,<sup>2</sup> or, less commonly, that the central idea of the spirited part of the soul is its reason-responsiveness, which characterizes some of our emotions and attitudes—such as anger (*thumos*, *orgê*), shame (*aischunê*), or love of honour (*philotimia*).<sup>3</sup>

This chapter takes up and develops the idea that seems distinctive of and importantly right about the latter account, which is that the spirited part is directly responsive to the reasoning part. Section 5.2 argues that in the *Republic*, the spirited part of the soul always follows the reasoning part, by acting on reason's settled judgements of good and bad when reason does not actively calculate, and being checked by and following reason's in-the-moment judgements when it does actively calculate. This



raises the question why we need a separate part of soul that always does reason's bidding, and whether, if the spirited part is entirely reason-following, Plato can nevertheless maintain its separateness from the reasoning part. Section 5.3 takes up the *Timaeus*' demiurgic perspective on the spirited part of the soul and argues that here, too, the main point about the spirited part of the soul is that it is reason's obedient servant, and that by playing this role it enables reason to delegate to it some of its work of managing embodied life, so reason can both rule effectively and contemplate in peace. Finally, Section 5.4 argues that the *Phaedrus* confirms this account of the role of the spirited part as reason's executor and uses it to foreground reason's philosophical nature.

## 5.2 The Spirited Part of the Soul in the *Republic*

The *Republic* characterizes the spirited part of the soul in four contexts. First, even before the argument for soul-division, in Book III, Socrates says that musical and gymnastic education are for the sake of the spirited (*to thumoeides*) and wisdom-loving elements in our nature, with the spirited element being the source of toughness, which becomes courageous if rightly nurtured (410b–411e). Second, in Book IV, Socrates argues that the spirited part is a distinct part of the soul; here, he makes some very general claims about the spirited part and considers some cases that are supposed to show it in action (440e–441c). Third, among the vicious characters Socrates describes in Books VIII and IX is the timocrat, who is ruled by the spirited part of the soul (548d–550b). Fourth, throughout the *Republic* Socrates makes various comments about anger, shame, the military, dogs, honour, victory, and so on, all associated somehow or other with the spirited part of the soul. The present account takes Socrates' general statements about the spirited part of the soul in Book IV as a constraint on the interpretation of spirited attitudes in the Book II–III account of pre-rational education, the Book IV conflict cases, the Book VIII spirit-ruled timocrat, and the references to spirited attitudes and tendencies elsewhere in the *Republic*.

To show the independence of the spirited kind in the soul, Socrates applies the principle of opposites to a story about Leontius, who desired to look at some corpses but also felt angry about doing so, and after some struggle, opened his eyes wide and simultaneously cursed them or his appetites for looking (439e–440a). Socrates says that Leontius' case establishes

that anger sometimes makes war against appetite, and that often when appetite forces someone contrary to rational calculation, that person reproaches himself and gets angry with that in him that's doing the forcing (440a–b). But then he makes a much stronger, universal, claim:

You can't say you've ever (*ouk an...pote*) seen spirit cooperate (*koinônêsanta*) with the appetites to do what reason has decided must not be done... When a person thinks he has done something unjust, if he's noble he won't resent bodily suffering if he thinks it is just punishment, but if he thinks he's suffering unjustly, his spirited part will become angry and fight back, enduring appetitive pains until it wins, dies, or is called to heel by reason. (440b5–d3)

Spirit isn't opposed to appetite when the reasoning part judges appetite's complaint justified, and in such a case spirit fights on appetite's behalf. But the spirited part of a noble person who thinks appetite's complaint unjustified fights against appetite. This raises the question: if it's the noble person's spirited part that opposes appetite when appetite is contrary to reason, what about the base person? Socrates' universal formulation ('You can't say you've ever seen spirit cooperate with the appetites to do what reason has decided must not be done') implies that even in a base person, spirit never opposes the reasoning part. Yet most readers take Socrates to qualify this claim in what follows, when he says that the spirited part is by nature an auxiliary of the reasoning part unless it has been corrupted by a bad upbringing (*epikouron on tòi logistikôi phusei ean mê hupo kakês trophês diaphtharêi*, 441a2–3). Most readers suppose that when Socrates says, 'by nature an auxiliary of reason... unless corrupted', he restricts the spirited part's unfailing obedience to reason to noble people. On this interpretation what Socrates means to say is that while by nature, that is, in a soul in its good condition, the spirited part sides with the reasoning part, it can, in a corrupted or poorly educated individual, be in the contrary-to-nature condition of siding with the appetite against the reasoning part.

But if this is right, why is Socrates so sure that Glaucon has never seen spirit oppose reason on the side of appetite? I propose that rather than qualifying the universal claim 'you've never seen spirit cooperate with the appetites to do what reason has decided must not be done', if we think about what an auxiliary (*epikouron*) is, we can see how 'spirit is the natural auxiliary of reason' might be building on it. An auxiliary in the political system of the *Republic* is an assistant or subordinate among

the guardians, one who performs governmental functions subordinate to and supportive of the rulers' functions. (By contrast, an 'ally'—as *epikouros* is often translated—might be an equal or superior.) Since it's rulers that have auxiliaries, the point is that because the spirited part always follows and never opposes the reasoning part, it is naturally suited to be reason's auxiliary—that is to say, *when* the reasoning part is the ruler, to be its helper (to which we might want to add, reason is a natural ruler (444d)). In other words, the spirited part is a natural auxiliary to reason (when reason is ruler) because it always sides with reason. Similarly, we might say of a child with a certain build and physical skill set, 'He's a natural quarterback', even though he doesn't know how to play American football, even if he will never play American football because, for example, he does not live in the United States. We may also treat the anger of non-human animals in the same way: were it to have the opportunity to follow reason, it would—for which domesticated animals might be thought to provide evidence. Socrates compares the spirited part of the soul to a dog, which, being a friend to what it knows and an enemy to what it doesn't, is 'by nature wisdom-loving' (375e–376b), and obedient to its human master (440d). These comparisons suggest that since dogs are like the spirited part by being spirited themselves, the spirited part of some non-rational animals is a natural auxiliary of something rational (the human), even if this is outside themselves.

Given this characterization of the spirited part of the soul, we can see why Socrates raises the worry whether the spirited part is independent enough of the reasoning part for the principle of opposites to ever apply. Glaucon's assurance that babies and non-human animals are angry (441b) does not tell for or against its independence from reason because these are cases in which (internal) reason couldn't rule, since it is not present. One could argue that even the anger of babies is a natural auxiliary of reason because once children begin acquiring beliefs about how to live, and then when they begin reasoning, their anger follows their reason, and it is because of the reason-following potential of anger that anger can be recruited to aid in reason's rule.

If the spirited part is indeed always obedient to reason as I've been arguing, then Socrates needs to give a very carefully constructed case of opposition to show that the part that has calculated is different from the part that is angry without calculation (441b–c). On the one hand, there must be opposition enough to establish two separate parts; on the other

hand, it can't be the kind of opposition that violates the principle that the spirited part always follows reasoning. Socrates' case is of Odysseus restraining his angry impulse to kill his maidservants until it's safer to do so. Here reasoning opposes spirit without spirit opposing it in return. At reason's chiding, spirit is 'called to heel' like a shepherd's dog.

Aristotle's remarks about *akrasia* due to anger help us understand the case: anger is like a hasty servant who runs off to carry out instructions without having heard them through to the end, and so does the wrong thing (*NE* vii.6 1149a25–30). This servant does not disobey any instructions he grasps, but his grasp of the instructions may be inadequate to carrying them out. This is presumably due to two deficiencies in the servant, the first of which cannot be corrected, while the other can. The first is that he cannot deliberate himself, and that is just a limitation of his nature. But the second is that he has rushed off without listening to the instructions in full, and that is something that can be corrected. A good servant waits on his master's orders. Still, a good servant also doesn't always require orders, because he knows his master's mind.

So although the spirited part never opposes the reasoning part, it is possible for the reasoning part to oppose the spirited part, on the basis of new information or new calculations, and when this happens, the spirited part adjusts in response. So we should think of the spirited part as having had its desires formed on the basis of the reasoning part's prior or standing judgements about what is good or best to do, but as updating its desires in response to new judgements by the reasoning part. (Further evidence of this is Socrates' claim that dialectic has destructive effects on the values of the young (537d–539a): they cease to honour the convictions about the fine and good that they have been brought up with once these have been undermined by argument; this is if honouring is an attitude of the spirited part of the soul.)

It might be objected that according to Socrates, we call an individual courageous when his spirited part preserves reason's pronouncements as to what is to be feared and what is not (442b–c), but on this account of the relationship between the spirited and reasoning parts, the spirited part always does this. But it can't be that the spirited part's preservation of the reasoning part's judgement is sufficient for courage no matter what the content of the reasoning part's judgement. In context, Socrates has just said that an individual is wise because of the (wisdom of the) smallest part in him. He continues: and what's responsible for his being

courageous is the spirited part—assuming, apparently, that we are talking about the same, wise, individual. It is a condition of the spirited part *in addition to the wisdom in the reasoning part* that makes a person courageous, and that condition is, ‘preservation of reason’s pronouncements’. Socrates’ claim that an individual is courageous when the spirited part preserves reason’s judgements about what is to be feared and what is not concerns *preserving* rather than *acting on* judgements about what is to be feared or not. The acting seems to be taken for granted. But Socrates envisions people ceasing to hold on to their judgements under the pressure of fear, pain, and so on. The spirited part contributes to courage the endurance or toughness that enables people not to change their judgements under ‘bewitchment’ of pleasures and pains (413b–e).<sup>4</sup>

It might also be objected that if there is never any possibility of the spirited part disobeying reason, the musical and gymnastic education of *Republic* II–III becomes redundant. But this objection assumes that the point of gymnastic and musical education is to make an initially recalcitrant spirited part obedient whereas on the present account it is to improve the spirited part’s sensitivity to reason’s command—that is, to make it a more careful listener. At the end of his account of the musical and physical education, Socrates says that music and gymnastics are not, as people think, for the sake of the soul and body, respectively, but for the sake of the soul (410b–c, 411e). He explains: the guardians have in their souls both a spirited and a philosophical nature, and a purely physical education would overstrain their spirited part and make them wilder (*agrioterōi*, 410d–e). A purely musical education, on the other hand, would soften and dissolve the spirited part. In the case of music, the mechanism seems to have to do with the spirited nature’s response to qualities of sound (411a–b), such as melody and rhythm, which are supposed to follow the words (398d). The mechanism of gymnastic education becomes clearer after Socrates has distinguished the parts of the soul: ‘the mixture of music and poetry makes them [viz. the reasoning and spirited parts, *logistikōn* and *thumoeidēs*] harmonize, tensing and nurturing the one by means of fine speeches and learning, and slackening and exhorting the other, making it gentle by means of harmony and rhythm’ (441e). This indicates that when it comes to the musical education, the reasoning part is educated by the words (speeches, stories, learning), whereas the spirited part is educated by the harmony and rhythm. ‘Slackening and exhorting’ the spirited part of the soul is, I suggest,

training it to be flexible in response to the content of reason's judgements. Even if the spirited part of the soul never disobeys the reasoning part, poorly trained anger or resistance can be too indiscriminate or too precipitate, and in that case the spirited part will be, although reason's auxiliary, not a particularly effective auxiliary.

This suggests that the advantage the spirited part of the soul brings is not speed, as in contemporary psychology's automatic processing, but toughness.<sup>5</sup> We might compare the *Laws*' idea that we are like puppets pulled by cords made of many metals, of which 'the golden cord of calculation' is pliant, but, as it exerts little force on its own, requires us to make the cords of other metals pull along with it (644e–645a). The relatively small reasoning part of the soul in the *Republic* is well served by having a tough auxiliary to which it can delegate tasks. But toughness and receptivity or flexibility are in some tension, and the point of early education is to produce the optimal combination.

The description of the timocratic character in *Republic* VIII displays how the spirited part of the soul acts and is affected when it is ruling. Ruling is especially awkward for the spirited part, since it does not begin to rule because its own desires—for control, victory, and high repute (581a–b)—are competing to take over the soul. Instead, the spirited part comes to rule in the young timocrat as a result of a power vacuum created by competing pulls, on the one side from his father's virtuous influence, and on the other from his mother's and servants' badgering him about the financial losses and insults they all suffer as a result of his father's virtue (549c–550b). While some of his mother's and servants' complaints are about honour (his father doesn't retaliate when insulted; he doesn't take revenge; he's not honoured and praised while the political meddlers are), the major forces pulling are virtue and wealth. Contrast the oligarch, democrat, and tyrant, the coming-to-be of each of whom begins at the instigation of the desires that end up ruling them (553a–d, 559d–561d, 571a–573b).

The example of the timocratic character may seem to suggest an alternative to my claim that the spirited part is a natural auxiliary of reason because it always defers to reason, so that when reason rules it is reason's auxiliary. It may seem that the explanation for the timocratic character being obedient to rulers must be that he is a lover of ruling and honour (549a) rather than that he is a lover of reason. And if what is true of the timocratic character is also true of the spirited part, no matter whether it is

in a virtuous or vicious soul, then it would seem to be possible for the spirited part to side with appetite against reason, for example in an appetitively ruled person. However, suppose it's the ruler, rather than reason, that the spirited part obeys. Even so, the spirited part doesn't seem to have any opportunity to oppose the reasoning part, for in the oligarch the appetitive part's rule involves the enslavement of the reasoning as well as the spirited part (553c–d). If the language of slavery is supposed to suggest that the higher parts are forced, the spirited part is not the appetitive part's *auxiliary*; if it is supposed to suggest that the higher parts adopt the ends of the ruling part as their own, the spirited part is not siding with appetite *against reason* because reason too is siding with appetite.

Finally, understanding the spirited part of the soul as the home of the reason-following attitudes can explain the fact that *Republic* X divides the soul into two (reasoning and reason-following vs. reason-opposing or reason-ignoring) rather than three parts. It seems odd that, after arguing that the soul is tripartite, and basing his account of virtue and vice on this tripartition, Plato should suddenly treat the soul as bipartite and leave it unclear whether the reason-opposed element in the soul is appetitive or spirited. But if the spirited part is always reason-following, the reason-opposing element must be appetitive, and Plato has no need to introduce a third part in the context of the argument of Book X.

Near the end of the *Republic*, Socrates says that so far he has just considered the soul in its embodied condition; to know the soul in its true nature, one should study it by means of reasoning (*logismôî*, 611c3), looking at its love of wisdom (*philosophia*, 611e1). In 'The Truth of Tripartition', Myles Burnyeat proposes that the *Timaeus* and *Phaedrus* explore two possible answers to the true nature of the soul (simple or divided), for the *Timaeus* describes partition as a consequence of embodiment whereas the *Phaedrus* describes the disembodied soul as itself tripartite. Sections 5.3 and 5.4 develop Burnyeat's idea that the *Timaeus* and *Phaedrus* take up the *Republic* 611a–612a project of discovering whether the soul is in itself simple or divided, arguing that these dialogues explore two accounts not only of the disembodied as well as embodied soul, but more particularly, of the soul in its activity of philosophy. It turns out that our account of the spirited part of the soul helps us to distinguish the activity of philosophy from the soul's care of bodily things.

### 5.3 The Spirited Part of the Soul in the *Timaeus*

In Section 4.4, we had a glimpse of the *Timaeus*' account of the intelligence in the construction of the cosmos and its parts,—an account showing how they are made in the best possible way—when we looked at the lesser gods' creation of mortal soul-parts and their bodily organs to serve the end of the soul's eventual release from mortal embodiment and return to its companion star. In the same vein, Timaeus says that the (ultimate) good that comes of our having eyesight is that it allows us to observe the movements of the heavenly bodies, which gives us number, physics, and finally, philosophy. We can actually see the circles of intelligence in the movements of the heavenly bodies, and by studying them, become intelligent, and god-like, ourselves (47a–c). So in the *Timaeus*, the soul's philosophy would seem to be the understanding of the goodness and order of the world, an understanding possessed by the souls that animate the heavenly bodies. This suggests that the question, 'why is the soul tripartite?' should be investigated by considering how the spirited part of the soul in particular serves this end.

Timaeus says that when the lesser gods take up the task of embodying our souls, they build a mortal soul, containing terrible but necessary (*anangkaiā*, 69c9) affections—pleasure, pains, boldness and fear, anger, expectation—which they fuse together with sense-perception and desire.<sup>6</sup> He comments that the gods construct the mortal type of soul 'as necessary' (*anangkaiōs*, 69d5). In this they are trying to stain the rational soul as little as necessary (*mê pasa ên anangkê*, 69d7).<sup>7</sup> To achieve this end, as they build the human body to house the soul, they locate the mortal soul in the trunk, separated from the head by a neck; they divide the trunk into two sections, housing the superior spirited part nearer the head so it can listen to reason and along with reason restrain by force (*biai*, 70a5) the appetitive part for whenever it won't willingly obey (*peithesthai hekon etheloi*, 70a7) reason's orders and reasoning; they put the heart near the spirited part so that it can, through the blood vessels, ready all the parts of the body for action if the spirited part 'boils over'<sup>8</sup> upon learning that the appetites are suffering or doing something wrong. This is so that the reasoning part can be in charge (70b). The gods put the lungs near the heart to cool it down when overheated by anger or fear,<sup>9</sup> so that it labours less and is thereby better able to help spirit serve reason (perhaps because the heating up/pounding is not useful to/is in excess of the sensitization); they put the appetitive part of the soul—which has appetites for things the



body needs and is necessary if there are going to be mortal animals—in the lower part of the trunk, tying it down here to ‘live at its trough’—in order to keep it from disturbing the reasoning part; since the appetitive part is unable to understand reason’s orders, or even if it understands them, to take care (*melein*, 71a5), they construct the liver as a surface to be stamped by the power of thought (*hê dunamis tôn dianoêmatôn*) from intellect (*ek nou*<sup>10</sup>) which reflects back visible images.

In ‘The Spirited Part of the Soul and its Natural Object’, Tad Brennan proposes that the good served by the spirited part of the soul is binding together two elements in the soul that are too different to interact directly: appetites that are deaf to reason, and reasoning that is too weak to train and punish appetites.<sup>11</sup> Brennan asks from the design stance of the Demiurge, ‘What is the problem to which the spirited part of the soul is the solution?’ His answer is that embodiment requires appetitive souls, and since appetites are not self-limiting, they need to be managed and restrained by something external to them and something that uses force. These requirements are met by an honour-system that stipulates the terms on which goods are assigned. The spirited part is attached to honour, where honour is something projected onto appetitive objects or onto the honourable because good (which reason detects clearly and spirit only dimly, 109–11) by the prevailing honour-system. Spirit’s self-directed anger belongs to its secondary function, self-policing (118–20), for which it reimagines internal appetitive competition in social terms, because only an appetite is strong enough to fight an appetite (125–7).

Brennan’s projectivism about honour also explains the puzzling fact that in the *Republic*’s account of the coming-to-be of the timocratic character and constitution, the love of honour does no work in explaining the coming to be of the timocratic political constitution; instead, it is the love of money and the bronze and iron types pulling against the love of virtue and the gold and silver types that causes the decline that ends in a timocracy (547b–c). Socrates’ surprising claim that although this constitution is mixed, the predominance of the spirited element makes the love of honour and victory most manifest in it (548b) is explained if we suppose that we have here an honour-system for the distribution of appetitive goods, for the timocrat, like the oligarch whom he resembles and into whom he quickly declines, is driven by the love of money (113–14).

However, the connection between the spirited part and honour that is so strong in the timocratic character and constitution in the *Republic* is absent from the *Timaeus*, which associates the spirited part with anger at injustice, whether one's own or someone else's against one.<sup>12</sup> Nor is there a general interaction problem between reason and the appetites that requires spirit's mediation. For example, when the power of thought projects images of pleasure and pain onto the liver, appetite sometimes obeys (71a–b), and in this case the spirited part doesn't seem to be involved except in the sensitizing of the body for action.<sup>13</sup> The spirited part needs to come in with force only against a disobedient appetite.

It may seem surprising that *Timaeus* treats this as a case of *persuasion*, and contrasts it with the force that the spirited part uses on appetite—especially since in the *Republic* the oligarch was said to curb his worse appetites by force and fear rather than persuasion (554c). Perhaps the idea is that providing the appetitive part with the only incentives and disincentives that belong to its nature to recognize and value is persuasion rather than force, but since a human being has by nature the ability to rationally persuade himself, if he instead motivates himself by fear, what he is using, relative to his nature, is force. So a case of persuading appetite may at the same time be a case of compelling the person.

If, according to the *Timaeus*, we need a spirited part of the soul to do just the thing the *Republic* says (on the account offered in Section 5.2)—to be reason's executor and strongman, and if the *Timaeus* shows us how, in our case, it is able to do this because of the design of the lesser gods in putting it together and locating it where they did—we are in a position to answer the question, 'What good does it do us to have a spirited soul-part?' The *Timaeus* emphasizes that soul-division allows the reasoning part to reason in peace, suggesting that one good of having such an auxiliary is that it allows the reasoning part to delegate some of its cares to the spirited part. Even if the *Timaeus'* spirited part needs the reasoning part to tell it that some injustice has been committed (and this may be a difference from the more personified spirited part in the *Republic*), it does not need the reasoning part to tell it how to respond in each case.

*Timaeus* is not at all clear how the spirited part receives reason's reports and commands. This is especially striking because he seems to

characterize the cognition of the appetitive part in some detail. Timaeus says that the appetitive kind of soul is able to perceive but not believe (77b–c), and this is likely related to the argument in the *Theaetetus* that perception cannot grasp those nonsensible properties (e.g. being, sameness, difference, number) that belong properly to reasoning (184e–186d), which Timaeus reflects in the composition of our rational souls by those very nonsensible properties.<sup>14</sup> The fact that intellect influences the appetitive soul by means of images recalls the *Philebus*' description of our soul as containing a scribe, which writes statements in our soul, and a painter, which renders the content of these statements in images (38e–39c). If we come with these thoughts to the spirited part of the soul, then it would seem that being mortal, the spirited part of the soul should also lack the power to believe: it is no more composed of being, sameness, and difference than the appetitive part. In this case, it would receive reason's reports (e.g. the report of injustice that makes spirit's power boil over and sensitize the limbs so they are ready for action) and orders in the form of images (e.g. a representation of a response to such an injustice in the past, or of the response anticipated). Despite its cognitive deficiency, it would always act on (what it represents to itself of) reason's orders.<sup>15</sup>

It may be objected: if the spirited part is just reason's auxiliary, why is it not better conceived of as habit?<sup>16</sup> Why does Plato emphasize its concern with the just and unjust and shame and status in the *Republic*, and its anger at the unjust and shameful in the *Republic* and *Timaeus*? I think the answer to this is that Plato thinks that anger and shame are not only reason-responsive but also what we might call 'hot' motivations, and so have a strength that is needed to pull against motivations contrary to reason (cf. *Laws* 644d–e). These motivations contrary to reason might be appetitive desires to enjoy some ill-gotten bodily good, or they might be appetitive desires to enjoy bodily safety and refrain from injury-risking actions.

Let's turn, finally, to the *Phaedrus* image of the tripartite soul as consisting of a chariot pulled by two horses (in us, a good and bad horse, which are, respectively, the spirited and appetitive parts) and steered by a charioteer (reason). Here, even the gods' souls have a pair of horses—but the gods don't need to manage their appetites or distribute competitive appetitive goods amongst themselves. Yet evidently they need a spirited part of the soul.

## 5.4 The Spirited Part of the Soul in the *Phaedrus*

In the *Phaedrus* Socrates first divides the soul into two, then into three, in two opposed speeches, the first against love (*erôs*) and the second in praise of it. A comparative examination of both speeches promises to tell us more about what work Plato uses the spirited part of the soul to do.<sup>17</sup>

Socrates' speech against love (237b–241d), given in response to Phaedrus' challenge that he compose a speech rivalling Lysias' speech about why it's better to be in a relationship with a non-lover than a lover, begins with a definition: love is desire (*epithumia*), but since even non-lovers desire beautiful things (237d), defining love requires understanding the elements in the soul that rule and lead us, namely the inborn (*emphuton*) desire for pleasures, and acquired (*epiktêtos*) judgement, which pursues what is best (*doxa ephimenê tou aristou*). This is enough to define the conditions relevant to saying what love is and why it is so bad. Moderation is when judgement leads towards the best with/by means of reason (*logôi*) and is master with/by means of power (*tôi kratei*); *hubris* is when appetite irrationally drags us towards pleasures and rules in us (237e–238a). Love is a species of *hubris*:

when, without reason, desire masters the judgement that is impelled towards the right, [and] is driven to pleasure in beauty, and under the power of kindred desires rushes forcefully towards the beauty of bodies [and] conquers [its] education (from which rushing it takes its name), it is called *erôs* (238b7–c4).

Even though it can happen that the two elements in the soul, desire for pleasure and judgement, agree (237e), this definition counts all love akratic—seeming not to countenance any desire for pleasures in the beauty of bodies approved by judgement. From the definition follow some characteristics of the lover that make a relationship with him bad for the boy: the lover wants to deprive the boy of any goods of soul,<sup>18</sup> body, and possessions that would enable him to resist the lover's taking pleasure in him (239a–240a); the lover is not faithful, for when intelligence and moderation re-establish their rule in his soul, he is ashamed and runs away (241a).

This account of love would seem to hold for the bad kind of love in *Republic*, in which *erôs* is the leading drone in the tyrant's soul (572e), but not of the correct love permitted in the city (in which the lover may

only touch the beloved as a son, for the sake of the fine, if he persuades him, because excessive pleasure, including sexual pleasure, has nothing in common with moderation, but shares in *hubris* and indiscipline (402d–403c)). But in the *Phaedrus*, instead of simply providing reassurance that the good kind of love comes with moderation, Socrates embraces the claims that the lover is mad rather than moderate (because acquired judgement is not in charge), and that the lover does mad things (from the standpoint of convention, as does the philosopher, 249d: ordinary people think the philosopher is mad, not realizing he's possessed by a god). When Socrates recants his first speech, he says that madness isn't simply bad; some madness comes from a god and so is the source of great goods for us (244a). Socrates shows how mad actions may be the result of a state of the soul different from the *hubris* described in his first speech, and superior not only to it, but also to the moderation described in the first speech. He does this to show the superiority of reason's best activity—philosophy—to reason's control over the appetites—moderation.

To see this, let's turn to Socrates' second speech, in praise of love (245c–256e). This also begins with an account of the soul. Some scholars say the first speech gives a popular account of the soul<sup>19</sup> and the second gives Plato's own. Be that as it may, our task is to see what it is about this second account that allows Socrates to express the truth about *erôs* being a divine and beneficial madness. Socrates begins this time by asserting that soul is immortal (245c5), for which he provides the following proof:

- (1) That which is always in motion is immortal (245c5).
- (2) Only a self-mover is always in motion (for it does not leave off being itself) (245c7–8).<sup>20</sup>
- (3) The source of motion for other things must be a self-mover (for it cannot both be the source and have another source) (245c8–d1).
- (4) A source must be ungenerated.
- (5) What is ungenerated is imperishable (245d1–4).
- (6) Therefore a self-mover must be [imperishable, and so] immortal.
- (7) Being a self-mover is what it is to be soul (245e2–4).
  - a. For every body that is moved from without is without soul, but that which is moved from within is ensouled, since moving from within is the nature of soul (245e4–6).

This already one-ups the first speech's opening account of the soul because before saying anything about soul's components, it says what soul *is*: source of motion, self-mover, immortal. From this follows the function of soul: to care for all that lacks soul, to patrol the heavens (246c, cf. 247a). The gods are each assigned to a post; they do their work and also periodically feast on the vision of the forms outside heaven. (247a–b) Apart from the feasting on the vision of the forms, this is very close to the activity of the world-soul in the *Timaeus*. And that tells us what, in the *Phaedrus*, is the philosophy in which we are observing soul: perhaps observing the movements of the heavens but certainly contemplating the forms that are the models on which the world is based.

After the proofs about all soul Socrates speaks about the soul's form (*idea*), but less confidently: being merely human, he can only say what the soul is like, namely, the joint inborn power (*sumphutôi dunamei*) of a winged team of horses and their charioteer (246a).<sup>21</sup> By contrast with the first speech's account of soul, in which judgement was acquired rather than inborn, here reason or intellect, represented by the charioteer, is also inborn. Further, Socrates says that every kind or part (*eidos*) of the soul grows wings (251b7). There are, then, two kinds of motion, upward and forward, and if the horses have a monopoly on forward motion (with intellect only steering and restraining them) they don't have a monopoly on upward movement.<sup>22</sup> Since soul is both source of motion and has care of all that is unensouled, it seems likely that the gods' horses are driving the heavenly bodies in their courses.

Since the gods have divided souls, division obviously isn't a consequence of mortal embodiment; rather, mortal embodiment is due to the loss or weakness of wings. The explanation for this is that by contrast with the gods, whose horses are both good (246a), human beings have one good and one bad horse (246a–b, 253d–e). While the good horse is noble, a lover of honour (*timês*) with modesty and respect (*meta sophrosunês te kai aidous*), a friend of true opinion, driven by commands and reason alone (*keleumati monon kai logôi*), the bad horse needs a whip, and, if it is badly trained, it can keep our souls from rising up to the vision of forms (253e–254a). Failure to see the forms results in mortal embodiment and dependence on opinion (248a–b). The kind of human embodiment the soul undergoes depends on how much or little it saw, and/or whether it forgets or does wrong (248d–e); embodiment thereafter depends on its choice (249a–b).

This psychological sketch allows Socrates to give an account of love that displays how it is a gift of the gods: love is a divine madness, in which the beauty of the boy 'down here' reminds the lover of the form, and so causes the lover's soul to grow wings; that is, to aspire to a godlike condition (249d). The beloved's beauty affects the different parts of the soul differently: the charioteer is filled with desire (253e); the good horse remains obedient to the charioteer and compelled by respect (*aidoi biazomenos*) (253e–254a, cf. *Timaeus* 71a–d); the bad horse tries to get sex and eventually the charioteer and the good horse approach the boy, which occasions in the charioteer a memory of beauty (254a–b).<sup>23</sup> Struck by the memory of beauty, the charioteer pulls back on the reins (254c, 254e).<sup>24</sup> While it is too much to say that the bad horse teaches the person about the beautiful,<sup>25</sup> it is the bad horse and not the good one that is the cause of the lover's approaching the beloved and thence being reminded of beauty. The good horse seems only to follow—when the bad horse won't quit pleading for sex, the charioteer and the good horse give in; when the charioteer pulls on the reins, the good horse sits voluntarily out of shame (*hup' aischunês*) and astonishment (254b–d). But what 'only following' does is to translate reason's directives into motivations.

In closing, I would like to propose an intellectual target of Plato's distinction between a contemplating, reasoning part of the soul and another part of the soul that simply follows it. At the end of the *Phaedrus* (278e–279a), just after his famous subordination of written discourse to the discourse that lives in souls and that, being based on knowledge, is responsive to the souls of those addressed, Phaedrus abruptly brings up the philosophical prospects of the (dramatically) young Isocrates. Distinguishing reason's philosophical nature serves Plato's polemical purpose vis-à-vis Isocrates' conception of philosophy. According to Isocrates, philosophy does not have the great power some teachers boast (*Against the Sophists* 15.11), of making its students able to know the future and be successful in action and speech, but must be combined with natural aptitude and experience (14) for success (18). In later work, Isocrates adds that disciplines which don't guide our action and speech, for example astronomy, music, the theorizing of Parmenides or Empedocles et al., don't deserve the name 'philosophy'; hence,

...it is fitting for me to define and make clear to you philosophy, rightly considered: ... Since it is not in the nature of human beings to get that knowledge

having which we would know what is to be done or said,...I consider wise (*sophous*) those able to arrive at opinions (*tais doxais*) that are for the most part of what is best (*tou beltistou*), and I consider philosophers (*philosophous*) those who spend their time in those [studies] from which they will get this sort of intelligence (*phronêsin*) most quickly. (*Antidosis* 15.271)<sup>26</sup>

In these writings, Isocrates, like Socrates in his first speech in the *Phaedrus*, seems to hold that opinion (*doxa*) is the best thing in us, and to say that honesty about philosophy's power and value demands that we accept this. In *Antidosis* he also says that the 'philosophical' pursuit that gets to *doxa* is the composition of speeches worthy of honour, which praise noble causes. The noble exempla used in these speeches are supposed to influence actions, with the result that skilful thinking and acting come to those who are 'philosophically and philotimically' disposed to speeches (15.266–278).

Distinguishing a spirited part of the soul and assigning it the task of preserving and acting on *doxa* serves Plato's purpose of showing that philosophy is an activity of reason's truth-seeking (inquiring, contemplating) nature, distinct from its products, lawfulness and acquired opinion. True opinion as to what is best and conduct that adheres to it are valuable, but, the image of the soul in Socrates' speech in praise of love seems to say, even non-human animals can share in these. The distinctive good sought by human beings is that sought by the charioteer, knowledge of fundamental truths, for its own sake.

## Notes

1. Williams, 'The Analogy of City and Soul'.
2. See J. M. Cooper, 'Plato's Theory of Human Motivation,' *History of Philosophy Quarterly*, 1 (1984), 3–21; I also took this view in my 'Imperfect Virtue', *Ancient Philosophy*, 18 (1998), 315–39. M. F. Burnyeat says: '[S]tatus-concerns' [which motivate the spirited part] ... form a set of reactive attitudes ... that we have because we are social animals' ('The Truth of Tripartition', *Proceedings of the Aristotelian Society*, 106 (2006), 1–22 at 12). J. Wilberding says, 'The spirited part is essentially social and other-directed; it is wholly dependent on the views of others ... [But since] the *thumos* can view reason itself as a kind of individual whose opinions matter ... Plato proposes "a radical reprogramming of our aversion to shame. It should be triggered not by the fortuitous opinions of one's fellow citizens but by the reliable verdict of



- reason” (‘Plato’s Two Forms of Second-Best Morality’, *Philosophical Review*, 118 (2009), 351–74 at 366–7).
3. Irwin says that the spirited part’s desires are good-dependent—dependent on the rational part’s conception of good—but narrow, that is, for what is good for the agent, not for what is all-things-considered best for the agent (*Plato’s Ethics*, 212); R. Singpurwalla writes, ‘[S]pirit aims to be fine and honourable, ... [which] involves living up to your rational views about how you should behave, despite appetitive temptations to the contrary’ (‘Why Spirit is the Natural Ally of Reason’, *Oxford Studies in Ancient Philosophy*, 44 (2013), 41–65 at 42–3). Singpurwalla tries to accommodate the social by arguing that social views about how to behave come from facts about what is fine and good (55), but it seems highly variable to what degree they do. I also worry that the idea that spirit aims to be fine and honourable trades on a slide from spirit’s pursuing honour (*timê*) to its wanting to be honourable (*kalon*). Further, *Republic* 505d–e suggests that independent of its connection to the good our desire for the *kalon* may be satisfied by appearances.
  4. I owe this account of ‘preservation’ to J. Wilburn, ‘Courage and the Spirited Part of the Soul in Plato’s *Republic*’, *Philosopher’s Imprint*, 15 (2015), 1–21. It is possible to accept this account of preservation and of the spirited part’s contribution to courage without accepting Wilburn’s more general claim, which is that *Republic* psychology requires reason’s endorsement of every action.
  5. This is where my account of the spirited part diverges from Irwin’s. I follow Irwin’s view that ‘While reason shows us that some actions of type F are good and others are bad, emotion tends to focus on F-type actions in general, without the discrimination that results from reasoning. ... *The desirable condition* [viz. of the soul overall] *is not the one in which my reactions always wait on complete rational reflexion, but the condition in which my tendencies to immediate reactions have been formed by the right sort of rational reflexion, causing them to focus on the right features of situations. In this case the rational part has a regulative role, but it ought not to be giving specific advice about what to do in this situation.*’ However, I disagree with his claims that, ‘If spirited reactions are more or less right in a fair number of cases, then their *immediacy* gives them an advantage over rational reflexion in cases where explicit reflection would be inappropriate’ and that although ‘the spirited part relies on the agent’s conception of what is good and right, ... [t]his claim does not imply that the spirited part never conflicts with the rational part, or even that it never endorses an action that is endorsed by the appetitive part *and rejected by the rational part*’ (*Plato’s Ethics*, 212–13, my emphases).
  6. The Demiurge’s speech to our souls before embodiment (42a–b) says that we will experience sense-perception that arises from forceful affections; desire mixed with pleasure and pain; and fear, anger, and however many are consequent upon

them ('them' = sense-perception and desire). Are the gods doing anything with these passions to make some good come of them? Possibly, when they combine (*sunkerasamenoi*) desire and sense-perception with pleasure, pain, daring, fear, and some other passions, they are making these passions functional by combining them on the one hand with particular objects of perception, and on the other hand with a desire, a motivation to act.

7. In *Plato's Natural Philosophy: A Study of the Timaeus-Critias* (Cambridge, 2008), T. K. Johansen reads the references to necessity in this passage ('dread and necessary affections' and 'they put the mortal soul together as necessary') as cases of simple necessity (the necessity that, given a thing's nature, it will act and be affected in certain ways), rather than hypothetical necessity (the necessity of some means, given a good end to be achieved) (147). But Timaeus talks about the lesser gods housing (*proôikodomoun*), mixing (*sunkerasamenoi*), and putting together (*sunethesan*) the mortal soul—all of which look like designing activities, not just the effects of putting the soul in the body. Johansen, however, seems to restrict hypothetical necessity to the construction of the body, not to anything the gods may be doing to make the mortal soul-parts that will go in the body (149ff.). But perhaps we should not think of making the body and making the soul-parts as separate processes.
8. *Cratylus* 419d: *Thumos* gets its name from *thuseôs*, the seething and boiling of the soul.
9. Like *Timaeus* 70b, the Hippocratic *Peri Kardiês* sect 1 says that the heart is at the 'guard post' (*en phulakêi*); like *Timaeus* 70c, sect. 5 reports that the naturally cool lungs, cooled further by inhalation (and some drink, sect. 2), surround the heart and thereby discipline the uncontrolledness of its heat (*kolazei . . . tèn akrasiên tou thermou*).
10. Perhaps this is not just the faculty of intellect but the virtue of intelligence.
11. T. Brennan, 'The Spirited Part of the Soul and its Natural Object', in R. Barney, T. Brennan, and C. Brittain (eds), *Plato and the Divided Self* (Cambridge, 2012), 102–27 at 123.
12. The absence of honour in the *Timaeus* characterization of the spirited part may be seen as confirmation of Chapter 4's suggestion that the 'characteristic' desires and pleasures of each soul-part in *Republic* IX were the desires and pleasures of a person ruled by the relevant soul-part.
13. NB the *hina* clause at 71b3–d4 has the power of thought as its subject throughout, and the spirited part is not doing anything. On the issue of whether appetite has the cognitive resources to obey reason, Lorenz, *Brute*, 101, says Timaeus' solution is 'less than successful' because "Timaeus thinks of the appearances that reason generates as being external to the soul . . . [which calls for] a story about how appetite can see the pictures that reason paints . . . '.

14. Recent scholarship has improved our understanding of the cognitive limitations of appetite in the *Timaeus* (and likely in the later dialogues as a whole). Bobonich argues that the late dialogues' non-rational soul-parts are cognitively impoverished (*Utopia Recast*); Lorenz argues that the specific cognitive impoverishment is the denial of belief, which is replaced in the late dialogues by imagination (*Brute*).
15. J. Wilburn argues, on the basis of *Timaeus*' repeatedly calling the cognitive power belonging to the mortal soul-parts irrational sense-perception, that there is no difference in the cognitive capacities of spirit and appetite; their difference is that spirit is obedient but appetite is not. See 'The Spirited Part of the Soul in Plato's *Timaeus*', *Journal of the History of Philosophy*, 52 (2014), 627–52. Wilburn supposes that the spirited part is obedient to reason, unlike the appetitive, because of the receptivity of honour, its central concern, to different social norms.
16. This objection is due to Hakan Genc.
17. Nussbaum recommends comparing both speeches to understand what the *Phaedrus* says about *erôs*, philosophy, and the good life in "This Story isn't True": Madness, Reason, and Recantation in the *Phaedrus*' (ch. 7 in *Fragility of Goodness*, 200–33). I am following her recommendation to understand what the *Phaedrus* is saying, about the soul—although I come to different conclusions. Nussbaum finds the first speech to agree substantially with the *Republic's* conception of non-intellectual elements in the soul as good-independent and indeed blind, as naturally tending to excess when in control, and 'the *logistikón* as . . . necessary and sufficient for the apprehension of truth and for right choice' (206). Of course I do not think this a proper account of the non-reasoning elements in the *Republic*, but also, Socrates' first speech in the *Phaedrus* notably doesn't call the superior part of the soul the '*logistikón*' or mention 'apprehension of the truth'.
18. Including 'divine philosophy', 239b4, on which H. Yunis (*Plato: Phaedrus* (Cambridge, 2011) [= *Phaedrus*]) comments: 'As S. utters the term *philosophia* in this passage without content and in support of conventional values, it has an Isocratean resonance.' E. Asmis, "Psychagogia" in Plato's "Phaedrus" (*Illinois Classical Studies*, XI (1986), 153–72) lists, in addition to praise of philosophy, the further Isocratean features of the speech: the clear demarcation of subject-matter aided by definition, the claim that human beings are guided by either *doxa* or *epithumia*, the insistence on importance of educating soul, and the attention to rhythm and arrangement (rather than novelty) of subtopics in the speech. Asmis argues that by Socrates' first speech Plato means to show the superiority of Isocrates to Lysias, and by the second, Isocrates' inferiority to the true rhetoric, which depends on the speaker's knowledge of the subject-matter and of the listener's soul, described at the

- dialogue's end but also (according to Asmis) instantiated by Socrates' first speech (on which see Chapter 1).
19. R. Hackforth (trans. and comm.), *Plato's Phaedrus* (Cambridge, 1952), 41–2; Yunis, *Phaedrus*, 113–14.
  20. What about the case when it ceases to exist and so, without abandoning its nature, ceases to move? R. Bett observes that *Phaedo* 105d makes exactly the same illegitimate move: the soul always brings life, so does not admit of death, so . . . is immortal ('Immortality and the Nature of the Soul in the *Phaedrus*', *Phronesis*, 31 (1986), 1–26).
  21. On the trope of likeness, especially in the Tisias section of the *Phaedrus*, see J. Bryan, 'Eikos in Plato's *Phaedrus*', in V. Wohl (ed.), *Probabilities, Hypotheticals and Counterfactuals in Ancient Greek Thought* (Cambridge, 2014), 30–46.
  22. So it isn't right that intellect is 'a relatively impotent moving force' (Nussbaum, *Fragility of Goodness*, 214).
  23. S. Obdrzalek, 'Contemplation and Self-Mastery in Plato's *Phaedrus*' (*Oxford Studies in Ancient Philosophy*, 42 (2012) [= 'Contemplation and Self-Mastery'], 77–107), says this is what corrects reason's tendency to get so caught up in ruling that it forgets that it has another end, contemplation (99–100).
  24. G. Ferrari comments that here self-control is achieved not intentionally but as a side-effect of contemplation (*Listening to the Cicadas* (Cambridge, 1990), 189); Obdrzalek follows ('Contemplation and Self-Mastery', 96). (Cf. *Republic* 485d on the 'stream of desires'.) But this assumes that the charioteer pulls on the reins as a result of falling back in awe at the vision of beauty, whereas the text only says that he pulls on the reins at the same time as he falls back, and that seems to have something to do with the fact that the form of moderation is next to the form of beauty. Is the charioteer 'compelled to pull on the reins' (254c) by moderation? Further, his use of pain-inducing violence on the bad horse seems deliberate, as in the control of appetite at *Timaeus* 71a–d and *Laws* 644d and ff.
  25. As claimed by Nussbaum, *Fragility of Goodness*, 221. In Socrates' second speech, 'the non-intellectual elements are necessary sources of motivational energy . . . intellect alone is a relatively impotent moving force . . . The non-intellectual elements have an important guiding role to play in our aspiration towards understanding . . . [through? their] keen natural responsiveness to beauty . . . The role of emotion and appetite as guides is motivational . . . But it is also cognitive: for they give the whole person *information* as to where goodness and beauty are . . .' (214–15).
  26. Although the *Antidosis* was composed after the *Phaedrus*, its dismissal of theoretical disciplines from philosophy on the ground that they are not useful to speech and action is a recurrent theme in Isocrates' writing, e.g. *Helen* 5.

## 6

# Psychological Eudaemonism and Explanation

## 6.1 Introduction

This final chapter examines a consequence for psychological explanation of the diminished role played by psychological eudaemonism in the account of Plato's moral psychology offered in this book. To review that diminished role: Chapter 2 argued that the psychological eudaemonism that Socrates introduces in the *Protagoras*, according to which we always do what we believe is the best of the things we can do, is a psychological hypothesis that (a) if true, would show how virtue could be teachable as Protagoras claims it is, and (b) is argued for on the assumption of ethical hedonism, which is rejected in Plato's other dialogues. And Chapter 3 argued that rather than psychological eudaemonism, our natural desire for what is good is the basis for Plato's claims across the dialogues that vice and wrongdoing are unwilling—for these in fact inhibit the fulfilment of that natural desire. However, even though psychological eudaemonism isn't the cornerstone of Socrates' theorizing about the soul, and rejecting or retaining it isn't as important to Plato's psychological views as previous interpreters have maintained, it seems significant that we feel we understand an action when we see how it appears best to its agent (Section 2.3 on the *Protagoras*) and that in the divided soul the three soul-parts are represented as agents each pursuing the good under some more or less adequate conception (Section 4.3 on the *Republic*).

In the middle of the twentieth century, William Dray argued for a type of explanation appropriate to the human sciences distinct from the natural sciences (for the latter, Dray accepted Hempel's covering law model according to which event E is explained by showing how it falls under a general law L according to which in circumstances C events of type

E occur).<sup>1</sup> According to Dray, since human action has a ‘thought-side’, the historian who seeks to understand an action must re-enact it in his own mind, engaging in vicarious practical reasoning that makes clear the inevitability, given the agent’s beliefs, of what was done; this, Dray says, shows the action’s point or rationale. In the human sciences, to see the action’s rationale from the point of view of the agent is to explain the action.<sup>2</sup> It seems to be a widespread assumption in contemporary Plato scholarship that psychological eudaemonism in Socrates/Plato functions similarly to explain an action by showing it to result from the agent’s (true or false) beliefs about good and bad in combination with her (unvarying) desire for the good.<sup>3</sup>

However, this chapter argues that for Plato, psychological eudaemonism’s potential to explain actions and agents is limited, and, in the case where the agent’s beliefs are false, both rife with moral hazard and in need of supplementation. Section 6.2 argues that in his *Republic* X criticism of poetry, Socrates describes the ethical dangers of seeing things from another’s point of view when that point of view is populated by false beliefs about good and bad; these dangers also arise for explanations of actions based on false beliefs about value. Section 6.3 gives a close reading of a passage in the *Phaedo* usually taken to derive an account of teleological explanation from the assumption of psychological eudaemonism, and shows that the point is rather that in the case of *intelligent agency* the agent’s judgements of goodness explain her actions—from which it does not follow that in the case of unintelligent agency, an agent’s false judgements would also explain her actions. A parallel between Plato’s treatment of character in relation to intentional explanations and forms in relation to teleological explanations confirms this suggestion (Section 6.3.1). Finally, Section 6.4 argues that according to Plato, intelligence explains actions not because beliefs explain, but because goodness does.

## 6.2 *Republic* X on Seeing Things from the Agent’s Point of View

*Republic* X’s criticism of poetry reveals one problem with taking the agent’s point of view to understand his actions. Seeing this requires a quick review of that criticism. Socrates says that since he has elaborated

the division of the soul, he is able to say more precisely why mimetic poetry is bad for us: it is an insult to our thought (*lôbê... dianoias*, 595b5–6). Mimetic poets are received as teachers about goodness and the virtues, yet they lack knowledge or true belief about these things and only represent how they seem to an equally ignorant audience (602a–b, cf. 492b–493d); the poets are, in short, imitators of mere appearances of virtue (*mimêtas eidôlôn aretês*, 600e5). To make things worse, there is an irrational part of the soul that is susceptible to appearances (602e–603a), and mimetic poetry strengthens this part at the expense of the reasoning part (605b–606d). Presumably the strengthening of the irrational part is a matter of the agent's becoming increasingly likely, in belief-formation and action, to follow appearances and ignore or reject reasoning when it conflicts with these appearances. For example, while reasoning teaches that it is not appropriate to react to losses with extreme lamentation—the loss may not be worse overall; but even if it is, lamenting inhibits deliberation about the future (604b–c)—poetry depicts agents lamenting as if the loss is so great that lamenting is the appropriate response, and when we are at the theatre we publicly praise and pity the lamenting agent without restraint and with pleasure (606a–c). This attitude of enjoying pitying others suffering cannot be contained, and inevitably bleeds over into our attitude towards ourselves when we suffer: we take misfortunes hard, and we indulge our suffering, finding some pleasure in doing so. Thus the institution of poetry supports the 'rule' of the irrational part of the soul and undermines that of the reasoning part.

While commentary on these passages naturally gravitates towards the controversial and consequential claim that we can't sequester our responses to poetry but must carry them over into how we live our lives,<sup>4</sup> my interest in this passage lies in what it suggests would be Plato's response to the contemporary view that coming to see an action from the agent's point of view constitutes an advance in explanation or understanding. According to Socrates, poetry leads us to understand and sympathize with a character's reaction, in action and feeling, to his suffering, and in doing so, we 'follow it, sympathize with the hero, take his sufferings seriously' (605d), and share in his pain but also take some pleasure in doing so (606b). There seem to be two problems with our response. First, we actually enjoy pitying the sufferer. Socrates doesn't say much about what this pleasure is, but in the case of tragedy it may be indulgence of the viewer's own empathetic feeling (*pathos*),<sup>5</sup> whereas

in the case of comedy it would be pleasure in the suffering of the character, which the *Philebus* counts as 'malice' and 'injustice' (48b). Second, and more important, the hero is a poor judge of how bad his misfortune is (because the poet is a poor judge), but in pitying the hero, we temporarily take his or the poet's judgement as our standard for how much lamentation and pity are appropriate. It may be objected that we only pretend to take up his standard temporarily, rather than actually taking it up, but since what is (actually or fictionally) taken up is a standard of feeling rather than action, and we may find ourselves weeping along with the lamenting hero, it's not clear how psychologically distinct the attitudes are, and we may be unbeknownst to ourselves training ourselves to respond to real-life losses with lamentation. Perhaps, if we had an independent and stable assessment of the badness of the hero's misfortune, we could pity him just the right amount for the misfortune, and also pity him for overreacting to it (as we sometimes do in real life, for example when we respond to a young child's injury sympathetically but without matching her assessment of the magnitude of her suffering). But we don't have that standard, and when we consume poetry we expect the poet to provide the standard. And the particular standard provided by Homer, Aeschylus, Sophocles, and Euripides indicates that misfortunes can ruin a good man, depriving him not only of wealth, family, or life, but also of virtue. Perhaps the truth is that non-virtuous characters respond variably to good and bad fortune, and the poets misrepresent this truth by representing virtuous characters as doing so as well. The irrational part of the soul most susceptible to poetry, Socrates says, doesn't know what the greater or smaller are, but believes that the same thing is now great, now small (605b–c). So viewing the variable responses of a hero, the ignorant theatregoer thinks, 'He was once so great; now he's completely reduced', when the truth is that he was never so great. In sum, whereas seeing things from others' point of view might seem to many modern writers on literature and philosophy to be one of the morally improving results of consuming literature, Plato's concern with the truth leads him to emphasize that depending on who those others are, such activity may be a rich source of false beliefs.

This suggests a reason why assigning the lower soul-parts characters that allow us to dismiss desires originating from them is a good counterweight to representing them as agents seeking the good under some faulty conception (Section 4.3). For while we are able to



adopt the point of view of the spirited or appetitive parts of a soul and evaluate things in terms of a limited subset of the goods available to human beings, when we do so, we risk making that point of view our own.

Poetry doesn't just give us false evaluative beliefs but also false causal beliefs. For example, reading the *Iliad*, I'm likely to conclude, 'Achilles refused to fight with the Achaeans because Agamemnon injured him when he deprived him of his prizes.' In other words, I take Agamemnon's action, rather than Achilles' appearances, to be the cause of Achilles' actions, because I take Achilles' appearances to be veridical. This isn't only true of poetry, however: in seeing how an action appears from the agent's point of view one always runs the risk of accepting falsehoods, because there is nothing critical of their point of view built into one's explanatory stance.

None of this is to say that Plato is unaware that seeing things from another's point of view is useful for certain purposes. For instance, in the *Gorgias* Socrates tells Callicles that we are able to communicate with one another about our experiences because we have the same experiences, even if in relation to different objects (as Socrates feels love for Alcibiades and philosophy, and Callicles for Pyrilampes' son Demos and the Athenian *dêmos* (481c5–d5)). And in the *Timaeus*, the account of the cosmos takes the form of a reconstruction of the reasoning of the Demiurge who made the world as good as possible. But, I maintain, the former passage describes communication of a motivating feeling, not explanation, and the latter passages conceive of explanatoriness being conferred by goodness, not the agent's beliefs. Sections 6.3 and 6.4 will make the case for this last point in greater detail.

### 6.3 Explanation by Intelligence in the *Phaedo*

This section describes the difference between Dray's idea that an agent's beliefs and desires give a rational explanation of her actions and Plato's idea that an agent's intelligence explains her actions. Prior to his final argument for the immortality of the soul in the *Phaedo*, Socrates recounts how his early interest in questions about natural phenomena—their growth, with what part of ourselves we think, and so on—foundered on puzzles raised by previous natural philosophers' accounts.<sup>6</sup> One problem, Socrates complains, is that these natural philosophers invoke opposites as causes of the same thing. For example, the coming to be of two things is

said to be caused at one time by their coming together, and at another time by their being separated (97a–b).<sup>7</sup> He tells how, in the midst of his confusion, learning that Anaxagoras said that everything is caused by intelligence (*nous*<sup>8</sup>) gave him a ray of hope:

... [O]nce when I heard someone reading from a book of Anaxagoras and saying that *nous* is the orderer and cause of all things, I was pleased by this cause, and it seemed to me to be good that *nous* is the cause of all things, and I considered that if this is how it is, then it is ordering *nous* that orders all things and places each thing in the manner and place that is the best; then, if one should wish to find out the cause of each thing, in what way it comes to be and passes away and is, one should find out this about it, in what way it is best for it to be or to undergo or to do anything. From this argument it is fitting for a person to seek to know nothing else, about this and about other things, other than the best. (97b–d)

Socrates' appropriation of Anaxagoras has the form: if intelligence is the cause of all natural phenomena, then the explanation of natural phenomena should take the form of showing how they are for the best.<sup>9</sup> Unfortunately, as Socrates goes on to complain, the causes Anaxagoras actually cites (air, ether, water) don't explain natural phenomena by showing how they are as they are for the best. Although some commentators suppose the problem with Anaxagoras' citing air, water, and ether as causes is the contradictory causality that Socrates has just been complaining about, Socrates' complaint about Anaxagoras is more specific. Socrates illustrates the specific complaint by saying about the explanation of human actions:

Indeed it seemed to me to be most like if someone were to say that Socrates does all the things he does by *nous*, and then when trying to say the causes of the things I do, he should say first that I am sitting here now on account of these things, because my body is composed of bones and tendons, and the bones are solid and have joints separated from each other, and the tendons are such as to tense and slacken, ... having forgotten to say the true causes, that, since it seemed best to the Athenians to find me guilty, on account of this it actually seemed to me better to sit here, and more just to remain to suffer the penalty which they commanded. [T]hese tendons and bones would long ago be at Megara or Boeotia, carried by a judgement of the best, if I had not thought it to be more just and fine to suffer whatever penalty was ordered by the city before fleeing and running away like a slave. To call such things causes is very strange. If someone should say that without having such things, bones and tendons and however many other things I have, I would not be able to do the things that seemed best to me, he would say something true. *But [to say] that I do the things I do on account of these*

*things, and do these things by nous but not by the choice of what is best would be a very great heedlessness of the account.* For [it is] not being able to distinguish that what is the real cause is one thing and that without which the cause could not ever be a cause is another. . . . (98c–99b, my emphasis)

What is the relationship between Socrates' desiderata for explaining his own action and for explaining natural phenomena? Are the desiderata for explaining his own action the basis for explaining natural phenomena?<sup>10</sup> Is the explanation of his own action an illustration of what the explanation of natural phenomena should look like? Is the explanation of his own action an instance of the explanation of natural phenomena?

What Socrates says about explaining his own action is that *if* Socrates acts from intelligence, *then* what seemed best to him must explain what he did (98c, 99a). In the imagined explanation of his sitting in prison by the arrangement of his bones and sinews Socrates specifically objects to the conjunction of the claims that (i) he acts from intelligence and (ii) the causes of what he is doing are material. Presumably this is because when someone acts *from* or *by means of* intelligence (as opposed to intelligence acting on him, or his luckily happening to do what is, independently, the best thing) he does so by exercising his judgement. So Socrates' sitting in jail, if due to his acting with intelligence, must be due to Socrates' correct judgement that it is best to abide by the verdict of the Athenian people.

It may be objected that since Socrates' explanation also invokes the Athenians' judgement of what is better, and he surely isn't crediting the Athenians with intelligence, Socrates must be saying that all actions, not only intelligent actions, are to be explained by the agent's beliefs about what is best. This is of course a possible reading, but an alternative is that since the explanandum is Socrates' action (not the Athenians'), the Athenians' verdict is described in the terms that Socrates, acting with intelligence, uses to determine his course of action ('A did what seemed best to her' is not ruled out for explaining a foolish action, and if the explanandum were the Athenians' action or judgement, 'it seemed best to the Athenians to convict Socrates' might itself need to be explained).

Explanation by intelligence doesn't simply replace air and water, or bones and sinews, and so on. Rather, intelligence displaces these from pretending to the status of cause to occupying the place of necessary conditions, conditions without which the cause can't act as a cause (99b). In this new role, these factors are no longer problematically contradictory as when they were contenders for cause. So while it is contradictory

for Socrates' bones and sinews to be *causes* of his being in jail in Athens because his bones and sinews would also be causes of his being in Megara, if his intelligently judging it best to remain in jail is the cause of his being in jail, then his bones and sinews can be necessary conditions for his being in jail, even though, had he judged it best to be in Megara, they would have been necessary conditions for his being in Megara. The difference is made by the fact that intelligence/knowledge can, for different purposes or in different circumstances, use the same materials to produce something and its opposite, or opposite materials to produce the same thing: for example, a potter can use the same clay to make a tall and narrow-mouthed jar or a short and wide-mouthed jar, or different kinds of clay to make the same kind of jar. The context of intelligent use dispels the mystery around the apparent contradictoriness of material causes.

Although this passage does not say how to explain actions not done from intelligence, the *Phaedo* does explain how the body is a source of false beliefs, including beliefs about good and bad, so it is reasonable to suppose that the body is part of the explanation of at least some unintelligent actions. The body's needs fill the soul with desires of all kinds which distract it from its all-important pursuit of truth; further, sense-perceptions misrepresent reality (66b–c); in particular, the soul can mistake the intensity of pleasures and pains for truth about the way things are (83c–d). These passages suggest that an important way in which the body contributes to foolish actions is by causing false beliefs or inhibiting the development of true beliefs.<sup>11</sup> Does it imply that this is the only way? The *Phaedo*'s concern is with the soul's pursuit of truth and consequently with how the body hinders this pursuit by leading it to form false beliefs, not with how to explain actions, except incidentally in the illustration of how to flesh out explanation by intelligence. So its silence about whether bodily desires result in foolish actions without first causing false beliefs needn't commit Plato either way, either to all actions being caused by the agent's beliefs about good and bad, or to some actions being caused by factors other than beliefs about good and bad.

Still, it does seem that the explanation of an action as due to false beliefs needs to be supplemented by an explanation for the false beliefs themselves. To the *Phaedo*'s pointing in the direction of the body's influence the *Republic* and *Timaeus* add the influence of our constitutions and our upbringing. So, for example, although in the *Republic* Socrates says that the oligarchic character steals from widows and

orphans because in dealings with them, his (standing) belief that wealth is the ultimate good operates unchecked by caution about detection and punishment (554a–d), Plato seems to think his false belief does not provide a final explanation, for he has Socrates add a historical explanation of the false belief: as a young man, the oligarchic character sees his timocratic father dashed down from high office into poverty, exile, or death, and decides to secure himself by amassing wealth (553a–c). The genetic story shows how wealth could come to seem so good to someone even though it is not. Again, in the *Timaeus*, where folly (*anoia*) is the ‘disease of the soul’, one part of folly, madness, is caused by excessive pleasures and pains and sustained by bad political constitutions (86b–87b). If we ask ‘in what do such explanations terminate?’ the answer of the *Timaeus* seems to be, ‘in an account which shows the good that this bad thing served/was a side-effect of’: false beliefs are the consequence, via pleasures and pains (perhaps exacerbated by our upbringing) of our embodiment in mortal bodies, but embodiment was necessary for the completion of the world with all kinds of living things, which is a very great good indeed.

### 6.3.1 *Explanation by Forms and Characters*

After his disappointment with Anaxagoras, Socrates reports in the *Phaedo*, he abandoned teleological and intentional explanation in favour of formal causal explanation (we encountered this passage in Section 1.2’s discussion of Plato’s method of hypothesis as a case of dialectical dependence):

I began in this way, and hypothesizing on each occasion an argument that I judge to be strongest, I posit as true those things which seem to me to agree (*sumphōnein*) with it, both concerning cause and all other things, and as not true those which do not, . . . I hypothesize that there is a beautiful by itself and a good and large and all other such things. If you give me these and agree that these things exist, from them I hope to show you cause . . . It seems to me, if something else is beautiful other than the beautiful itself, it is not beautiful through another thing than because it participates in that beautiful [itself] . . . (100a–c)

Forms cause the properties of objects encountered in experience. How they do this Socrates does not know precisely (100d), but he says that the experienced objects have the properties they do in virtue of ‘participating’ or ‘sharing’ in the perfect exemplars of those properties, where the nature of these exemplars is such as to cause instances in their own image.

Socrates says that he clings to this kind of explanation or cause 'simply, naively and perhaps foolishly', for these are 'the safest' causes.

Formal causation is 'safe' because it saves us from positing contradictory causes. If we think, for example, that the cause of beauty is the juxtaposition of certain colours, we will eventually find out that while this juxtaposition of colours causes beauty on canvases, it causes ugliness in my clothes, and that a single colour all by itself also causes beauty, for example, in a rose; however, if we think that anything beautiful is beautiful by its relation to the beautiful-itself, we will entirely avoid such contradictions. On the other hand, formal causal explanation defers the substantive explanatory work of providing an account of beauty (which must cover the beauty of all beautiful things and no non-beautiful things), of determining the process or mechanism by which the form of beauty comes to be in sensible beauties, and of explaining why it does so in such-and-such circumstances and not others. For to say that sensible S is beautiful by participating in the form of beauty is to leave the participation itself as a brute fact. Socrates also makes a 'cleverer' (105c) use of forms: sensibles, insofar as they participate in forms, inherit the relations of contrariety that obtain between forms (for example, hot being the opposite of cold, hotness in a body recedes at the advance of coldness (102d-e)); further, some natural things depend for their existence on participation in certain forms, so that, for example, fire, which always 'brings with it' the hot, cannot admit of cold, even though cold is not the contrary of fire.

The virtues and vices are the same sort of causes to our actions as hotness or coldness are causes to natural things. Even though we lack accounts of temperance-itself and wisdom-itself and justice-itself, we can still say that the virtues exclude the vices, and are such as to cause temperate and wise and just actions, and not intemperate, foolish, and unjust actions. Further, Socrates also makes the 'cleverer' use of the theory in the explanation of actions, as when he argues in the *Protagoras* (see Section 2.3) from the facts that (a) acting foolishly is acting from folly and (b) acting temperately is acting from temperance and (c) acting foolishly is the opposite of acting temperately, that (d) folly is the opposite of temperance (which means that temperance must be wisdom, since a thing can have only one opposite, and the opposite of folly is wisdom) (332a-e). Here, we might say, it emerges that intemperance always 'brings with it' folly.

This brief review of formal explanations shows their inferiority to explanations by intelligence. Formal explanation cannot say *how* Socrates' bones and sinews are necessary conditions that contribute to Socrates' sitting in jail in Athens (and so a fortiori cannot resolve the contradictoriness of their role in both this and Socrates' being in Megara); explanation by intelligence can. For example, one might say, 'it's just to obey the laws and decrees of one's city; Socrates is just; the just person is such as to do just things and not unjust things; the decree of Athens is that Socrates should remain in jail.' It follows that Socrates will be in Athens in jail, along with his bones and sinews, but on this account the bones and sinews are not only not causes but also not contributors—as they are in the account that appeals to his decision to remain in Athens and face his death, for this decision requires those bones and sinews for its execution: his bones and sinews enable his walking to jail, and sitting there after taking his hemlock, and so on.

It may also be that formal causal explanation is incomplete. Formal causation is a case of like causing like, of which it may be possible to ask 'why?' That Plato may think this is suggested by a development in his theology. In the *Republic* Socrates explains why a god cannot be the cause of bad but only of good things by arguing that the god is good, that the power of goodness is to benefit, not to be harmful, thus not to do harm, thus not to cause anything bad (379d). But in the *Timaeus* the explanation is that God, being wholly good, has no malice (*phthonos*, pain in another's good fortune or pleasure in another's misfortune, cf. *Philebus* 48b), and so no reluctance to make the world as like himself, that is, as good, as possible (29e–30b). Between these texts, the treatment of a god as a power such as to cause only good things is translated into an intelligent craftsman's unhindered technical motivation to produce the best product of which his craft is capable, given the materials. The bulk of *Timaeus'* account is of how God, looking to the living thing itself (i.e. the form), makes the cosmos as good as possible. The more that is said in favour of God's making living things as he does, the less of a role seems to be left for the form, or the more we are inclined to think of the form as the craft-knowledge of cosmos-making that God has.<sup>12</sup>

Similarly in the human case, 'she  $\phi$ -ed because it was the virtuous thing to do' awaits amplification, which should take the form of showing what good  $\phi$ -ing brings about, and how. The asymmetry between good and bad seems present in both the cosmic and the human

cases: *nous* and virtue explain because they bring about good things, folly and vice do not.

## 6.4 Psychological and Teleological Explanations

The alternative reading of the *Phaedo* offered in Section 6.3 according to which Socrates is committed only to the necessity and final explanatoriness for action of an intelligent agent's beliefs about what's good, raises a natural question. Why should intelligence explain but folly not? Socrates' response to Anaxagoras' '*nous* rules' in the *Phaedo*, that if intelligence is the cause, then all one needs to do is to show how the effects are good, suggests the answer: goodness explains. In the case of the crafts, intelligence is the cause of good products peculiar to the craft (good shoes in the case of shoemaking, good laws in the case of legislation); the *Timaeus* supposes a demiurgic God and shows how various features of the physical world (of which we are a part) are good in some way, either being good themselves or serving some other good end. But it isn't as if folly and only folly is guaranteed to have the opposite effects, so that one could say of every non-good thing, 'that's due to folly!' For some non-good things are necessary consequences of the good things that are due to intelligence; for example, the disorderly motion of souls consequent upon their embodiment (43a–e) is a necessary consequence of the Demiurge's decision that the cosmos be as perfect as possible, containing all the kinds of living things whose originals are contained in the living thing itself (39e–40a).

Goodness explains not only products and actions but also desires and indeed all the soul's pursuits. To begin with the soul's desires, let's turn to Socrates' report of the priestess Diotima's teaching about love in the *Symposium*, which we may divide into three main points:

- (S1) Love is of something (call this 'the beloved'), which is something beautiful (this point is taken up from Agathon, 195a–196b), which the lover lacks (199d–201c); if beautiful things are good things, then love is of good things, and love aims to have good things. Having good things forever is happiness ('love's end') (201c–206a).
- (S2) Because of (S1), love's works (*erga*) are reproductive of what lovers value, or conceive to be good: reproduction is mortal



creatures' way of approximating (since they can't have it) eternal possession of the good (206b–209e).

(S3) Love's end (*telos*) is knowledge of the form of beauty (210a–212b).

(S1) is already a little different from psychological eudaemonism because the lover is attracted to something beautiful—I suppose *qua* beautiful—rather than something good or believed to be good *qua* good. When Diotima asks Socrates what the lover of beautiful things, who wants to possess these beautiful things, will have when he possesses them, Socrates is stymied. So she nudges him forward, suggesting he substitute 'good things' for 'beautiful things'—but this is a theorist's, not obviously a lover's, substitution. The substitution allows Socrates to answer.

[DIOTIMA:] “Say, Socrates, the lover of good things loves; what does he love?”

[SOCRATES:] That they come to be his.

[DIOTIMA:] And what will he have when they come to be his?

[SOCRATES:] This [question] is easier, and I am able to answer that he will be happy.

[DIOTIMA:] The happy are happy by possession of good things, and no longer need one ask “Why does the one who wishes to be happy wish it?” Rather, the answer seems to be final (*Symposium* 204e2–205a3).

The finality of ‘it’ll bring happiness’ is supposed to indicate that explanation has come to an end. But why is ‘it’ll bring happiness’ such a stopping point? On the conception of explanation as showing the rationale of the action from the agent’s point of view, it would be because the desire for happiness is a desire we all, agents and explainers, have in common. Diotima says that the desire for good things (*epithumia tôn agathôn*) is common to everyone (*tên boulêsin kai ton erôta touton . . . koinon einai pantôn anthrôpôn*) (205a–206a). However, stopping asking ‘why?’ when we come to a shared desire makes the explainer the standard of intelligibility, but surely this makes explanatory satisfaction highly subjective and variable. Further, such an understanding based on common experience would have to extend to all mortal creatures, whose reproductive behaviour is supposed to be explained by (S1), but on the basis of what would we do this, given the vast differences between our constitutions?

One might think that although the account of love extends in principle to all mortal creatures, Diotima's focus must be on human beings with whom we can share appearances. For, one might reason, the beloved specified in (S1) must be something that appears good whether it is in fact or not, given that Diotima's substitution of 'good things' for 'beautiful things' is only valid if 'good things' are things that are attractive; that is, they appear good.<sup>13</sup> Diotima's account emphasizes that another necessary condition for love, in addition to the lover's lack, is that the lover *recognizes* his lack: the one who doesn't *think* he is in need won't desire that which he thinks he doesn't need (204a6–7). This seems to suggest that appearing good is at least a necessary condition for something to be an object of love. (Socrates argues that it a necessary condition of desire that the desirer have some kind of cognitive contact—for example, in memory—with that which is desired at *Philebus* 35a–c.)

However, while appearing good seems to be a necessary condition of love, Diotima makes it clear at three points that it isn't a sufficient condition. First, she suggests that it is necessary not only that the lover believes he has a lack but also that the lack is real: it is impossible for someone who is, for example, big or strong to want (*bouloitoi*) to be big or strong (200b4–5). True, people who are healthy *say* 'I want healthy things and to be healthy', and 'I desire those things which I have', but since they necessarily are what they are in the present, and necessarily have what they have in the present, what they are *really saying* (*ei allo ti legeis ê*, 200d4–5) is, 'I want the health I now have to belong to me in the future' (200c–d). Second, she follows up the substitution of 'good things' for 'beautiful things' by explaining that the lover of good things wants to have them, because by having them he'll be happy. But it isn't by having apparently good things that one is happy. Third, she rejects Aristophanes' account of what lovers desire on the grounds that we do not seek what is our own, or to be whole, *unless it turns out to be good*—for human beings are even willing to cut off their limbs if they think these are harmful; we don't value our own unless we call the good our own, she says; human beings love nothing other than the good (205e–206a). With this, Diotima denies that thinking good is sufficient for loving and adds that the thing loved must also be good.<sup>14</sup>

In Chapter 3 we saw that in the *Gorgias* and *Meno*, Socrates deattributes desires on the basis of the badness of their objects for the desirer; in Chapter 4 we saw that the *Republic* corrects this move by

acknowledging that there are multiple types of desire, some of which, even if they are responsive to appearances of goodness, do not always track the truth about it, and so remain attributable to the agent even if their objects turn out not to be good. Diotima seems to be working with the pre-*Republic* view of desire as cancelled if its object isn't in fact good.

In (S2) Diotima turns to how lovers pursue (*diôkei*) this state, and what is love's work (*ergon*) (206b). The answer is that happiness is pursued in reproduction, which is a mortal creature's proxy for eternal possession of the good/happiness. There are two ways in which reproduction is a proxy: first, one's offspring are not the same as oneself, but are only one's descendants; second, the goodness of what one reproduces depends on what one is pregnant with, and in the case of human beings, what one is pregnant with is what one values, whether physical offspring, or laws and institutions, or discourses. This is why counted among human lovers of good things and happiness are athletes, poets, lawgivers, and so on. Finally, (S3) Diotima tells us what the state of successful love, love that has attained its end, is: knowledge of the form of beauty.

Whereas the *Gorgias* considers actions instrumentally, as means to some end (Section 3.4), this discussion in the *Symposium* considers actions expressively, in terms of what conception of the good they are inspired by. Although the two discussions have different emphases, they are consistent. Both distinguish the end (my happiness, rather than what I believe to be my happiness) from what I do in its pursuit (do/make what seems to me conducive to my happiness). In the *Symposium* account, the good/beautiful beloved inspires and guides action, and the works of love—biological offspring in the case of all animals, and products of craft, legislation, poetry, and so on in the case of human beings—are the beloved-guided proxies for the end, happiness. Human beings presumably believe these works to be good, and their beliefs may be true or false; but might not some creatures accomplish their works without any belief or even representation? Diotima points to a difference between human beings and other animals in that a human being could perform the works of love because of reasoning (*oioit' an tis ek logismou tauta poiein*), but this is not the case with the other animals (207b–c). So, for example, birds produce chicks and hunt for worms for them but without thinking, 'this is how I get to be happy!' or 'These chicks are my way of living on!' So what explains their behaviour? Diotima says that *love* by its nature aims at immortal possession of the good (207c–d), and that

reproduction is the device (*mêchanê*) by which mortal creatures participate in immortality (208b). It seems that the *Symposium* gives a teleological, rather than a psychological eudaemonist, account of the works of love, with human beings having the capacity (which they may or may not develop) to understand the teleology and base their actions on it.<sup>15</sup>

Lastly, a passage in the *Republic* we have encountered more than once may be better understood as giving a statement of teleology despite its sounding like a statement of psychological eudaemonism:

... [I]n the case of just and fine things, many people would choose (*helointo*) things that seem [*ta dokounta*, sc. to be just and fine], even if they are not, and to seem to (*dokein*) do and acquire these things. But in the case of good things, to acquire the things that seem (*ta dokounta*) satisfies (*arkei*) no-one, but everyone seeks the real things (*ta onta zêtousin*) and disdains mere belief (*tên doxan*) ... The good is what every soul pursues (*diôkei*), and for its sake (*toutou heneka*) does everything, divining (*apomanteuomenê*) that it is something but perplexed (*aporousa*) and unable to grasp adequately what it is or to attain any stable belief such as it has about other things. (505d5–e3)

Socrates says that pursuit of the good belongs to *every* soul (or perhaps all soul, soul as such), not just rational souls or human souls. There is some cognitive component to this pursuit: the soul ‘divines’ that the good is ‘something’; it ‘is perplexed’ or ‘puzzles’ about what it is. We might compare *Philebus* 20d8–9: ‘everything that knows (*to gignôskon*) it [the good] pursues it and goes after it, wanting (*boulomenon*) to take it and possess it for itself.’ This isn’t a high level of knowledge.<sup>16</sup> On the other hand, there is something investigative about the attitude. Let’s take each of these points in turn.

First, rather than saying that we only pursue what we have rationally determined is good, the passage suggests that even souls that only divine that the good is something pursue the good. The claim, ‘The good is what every soul pursues (*diôkei*), and for the sake of which it does everything (*toutou heneka panta prattein*)’ includes one element of psychological eudaemonism, which is the reference to action: it’s not only that we want the good, but that we ‘do everything’ to achieve it (‘do everything’ may mean ‘perform each and every action’ or ‘exert great effort’). But what about the other element, doing only what seems best? As we saw in Chapters 4 and 5, the *Republic* recognizes different soul-kinds, a reasoning, a spirited, and an appetitive kind, and talks as if all three kinds have not only appearances but also beliefs of their own. If we say that

each soul-part has its own beliefs then we might equate 'doing what seems best to part such-and-such' with 'being motivated by part such-and-such'.<sup>17</sup>

Second, our beliefs about goodness have an intrinsic propensity to change compared to our beliefs about what is just and fine. It is implausible that this should be because we are stably in possession of the truth about what is just and fine, but it is more plausible that everyone has a motivation to possess the truth about goodness in particular, which unsettles false or unsecured beliefs about goodness—whereas the motivation to possess the truth about the just and the fine is not universal. For every soul 'divines' that the good is 'something': in us, this amounts to its recognizing that 'good' is not just a conventional designation; Socrates contrasts every soul's pursuit of real goods with the conventionalism with which many of us are content when it comes to the just and fine. Many of us think that what is just and fine is defined by convention, and pursue those things that convention determines are just and fine; however, in the case of the good, we seek what is by nature good, and we will not be satisfied to pursue what is conventionally defined as good—unless of course we believe convention tells the truth about what is good by nature. Our reality-oriented desire for good makes our beliefs about good and bad particularly reality-responsive. Some of us may even actively seek an account of what is good. This means that evidence—empirical, argumentative—that one's conception of the good is false will motivate belief-change; hence the special instability of beliefs about what is good.

The *Republic* passage suggests a looser but more global connection between our actions and the good than does psychological eudaemonism: our natural desire for the good, even though we don't know its identity, explains the directedness of many things we do: most obviously, pursuing what we believe is good (when we do, and being the sorts of creatures we are, we can do this often), but also trying to find out what the good is, which shows up in our sometimes being dissatisfied even though we have what we believe(d) good.

If souls have a natural orientation to our own good, then by definition a good soul would be able to secure its own good—but its way of doing so would depend on what kind of soul it is. If it is a human soul, and the best way of securing the human good is by using reasoning or calculation, then it follows that a virtuous human being would be knowledgeable or intelligent in the relevant kind of reasoning or calculation. This,

I submit, is the sense of 'virtue is knowledge' to which Plato is committed—rather than a denial that non-rational motivations can (directly) motivate our actions.<sup>18</sup>

## Notes

1. W. Dray, *Laws and Explanation in History* (Oxford, 1957), 118–55; Carl Hempel, *Aspects of Scientific Explanation* (Macmillan, 1965), 231–43. It may seem that this debate was resolved by Donald Davidson's argument that reasons can be causes because mental events (e.g. believing, knowing, noticing) are token-identical with physical events, although the rational principles connecting mental events characterized intentionally are distinct from and independent of the causal laws connecting these same events characterized physically. See, e.g., Davidson (1963) 'Actions, Reasons, and Causes' and (1970) 'Mental Events', in D. Davidson, *Essays on Actions and Events* (Oxford, 1980), 3–20 and 207–24 respectively. However, Davidson's way of allowing for the compatibility of the two types of explanation does not really address the question of what it is to make human actions intelligible. I'm grateful to Ken Winkler for discussion about this issue.
2. Dray was bringing to English-language philosophy an idea of the distinctive understanding appropriate to the human sciences found in Dilthey, and elaborated in Weber. Thanks to Michelle Kosch for this observation.
3. Cf. Irwin, *Plato's Ethics*, 53–5, 210; Penner, 'Socrates on the Strength of Knowledge' and 'Desire and Power in Socrates'; Penner and Rowe, 'Desire for Good'; R. Barney, 'Plato on the Desire for the Good', in S. Tenenbaum (ed.), *Desire, Practical Reason, and the Good* (Oxford, 2010), 34–64. In 'Psychological Eudaemonism and Interpretation in Greek Ethics', M. LeBar and N. Goldberg argue that psychological eudaemonism is the result of two requirements of interpretation: (1) understanding oneself as practically rational requires attributing to oneself the desire to live well and (2) interpreting another as practically rational requires attributing to them intentions one would have in their circumstances (in R. Kamtekar (ed.), *Virtue and Happiness: Essays in Honour of Julia Annas, Oxford Studies in Ancient Philosophy Supplementary Volume* (2012), 287–319).
4. In a relatively recent treatment of the issue, J. Harold, 'Infected by Evil' (*Philosophical Explorations*, 8 (2005), 173–87), uses a descendant of Plato's cognitive division of the soul, viz. of 'controlled' vs. 'automatic' psychological processes, to discuss the effects of identifying with bad fictional characters on our automatic processes. Thanks to Shaun Nichols for pointing me to this discussion.
5. T. Gould, *The Ancient Quarrel Between Poetry and Philosophy* (Princeton, 1990).

6. There is a widespread view that the *Phaedo*'s talk of forms and its interests in natural philosophy indicate that the *Phaedo* must be a middle-period dialogue reflecting the views and interests of Plato rather than Socrates—despite the evidence of stylometry that the dialogue is early. The justification for this is presumably Aristotle's statement in the *Metaphysics* that Socrates concerned himself with ethical matters but not with nature as a whole (987b1–2), but Xenophon's *Memorabilia* also shows Socrates to be concerned with questions of natural philosophy and teleology (I.4). Perhaps Aristotle intends to contrast ethical matters, which (prior to Aristotle's division of knowledge into practical, productive, and theoretical) would be thought of as only pertaining to human nature, with the study of nature as a whole—at which the Socrates of the *Phaedo* too confesses himself at a loss. (G. Betegh has argued that the account that the *Phaedo*'s Socrates says he is unable to provide is specifically a 'tale', a narrative of how the current state of affairs in the physical world is an intelligent solution to an original problem—for Socrates is able to give a teleological account of how things that are already in existence function for the best. See 'Tale, Theology and Teleology in Plato's *Phaedo*', in C. Partenie (ed.), *Plato's Myths* (Cambridge, 2011), 77–100. This would remove the apparent inconsistency between Socrates' claim that the soul's afterlife is for the best and his claim that he does not have the sort of account he hoped for from Anaxagoras. But we can also explain Socrates' 'for the best' accounts of the afterlife in the absence of teleological knowledge by his caveats about the merely likely status of these accounts). In any case, any account of Plato's Socrates needs to accommodate the *Phaedo*'s report of his early interest in natural philosophy and the *Parmenides*' account of his early conception of forms as objects of thought and knowledge.
7. D. Sedley, 'Platonic Causes' (*Phronesis*, 43 (1998), 114–32) summarizes Plato's standard for non-contradictory explanation as follows: cause F cannot be the cause of anything's being not-F, nor can F be the cause of G and not-G, nor can F and not-F be the cause of G. According to Sedley, this standard is met by both formal causation and intelligence, the latter as a case of good being the cause of good. But there's more to intelligence-explanations than to formal causal ones.
8. I will leave *nous* untranslated in quotations but translate by 'intelligence' in discussion. For *nous* as a virtue, see S. Menn, *Plato on God as Nous* (Carbondale, 1995). I'm also indebted to Stephen Menn for discussion of this passage.
9. In translation and discussion I am using the English 'cause' for Plato's *hê aitia/aitios/to aition* (what is responsible for). By 'explanation' I mean what we should say about causes—I don't accept the view of G. Vlastos ('Reasons and Causes in the *Phaedo*', *Philosophical Review*, 78 (1969), 291–325) and those following that the *Phaedo* is concerned with explanation more broadly rather than causation alone, and that Formal causes give logical explanations and intelligence (*nous*) gives reasons.

10. For example, D. Bostock says about this passage that Socrates is hypothesizing a divine mind that, by analogy with a human mind, has purposes or reasons for acting, on the 'over-optimistic' assumption that 'a man's purpose will always be to achieve what seems best to him' (D. Bostock, *Plato's Phaedo* (Oxford, 1986), 142–3). On my view, by contrast, Socrates' model of explanation here is not what he offered at the end of the *Protagoras* (it's human nature always to go after what one believes is best, no matter whether one knows what's best or not, and no matter what other forces are in the mind); rather, the model is of craft-knowledge, intelligence in the production of good things. Given intelligence, purpose is explanatory; nothing follows about what is explanatory in the absence of intelligence.
11. In *Socratic Moral Psychology*, Brickhouse and Smith claim this is the view of the 'Socratic' dialogues but suggest that the *Phaedo* has non-rational motivations that directly produce actions, without reason's involvement (197–9). This is odd given that the *Phaedo* provides many more examples of non-rational motivations influencing beliefs than do the 'Socratic' dialogues.
12. S. Broadie, *Nature and Divinity in Plato's Timaeus* (Cambridge, 2012), argues that the living thing itself functions as a sort of recipe that the Demiurge uses in creating the cosmos.
13. Anscombe, often invoked for the idea that all desire is under the guise of the good (see *Intention*, s. 40, 75f.), says, 'But the good (perhaps falsely) conceived by the agent to characterize the thing [desired] must *really* be one of the many forms of good' (76–7). I can't claim to have explained anything by saying (Anscombe's example), "She wants a saucer of mud because that appears good to her." But a little later, Anscombe says, an explanation of this form satisfies only when I have shown how the saucer of mud seems (correctly or incorrectly) to be really good: useful (I can sprout the beans in it), or beautiful (I can complete my shocking still life with it), etc. If Anscombe's point is that I can be wrong about the actual goodness of the object of my desire, but not in my conception of the good—if the explanation is to satisfy after all—then her position is very like Dray's. On the other hand, if the point is that what explains is (real) goodness (whether pleasantness or beauty or healthiness, etc.) of my object, then her position is more like Plato's.
14. Cf. *Republic* 334b–335a, where supposing that the friend is the one believed to be good has the unacceptable implication that, in the case where one's belief is mistaken, it is just to harm good people and benefit bad ones. An anonymous referee complains that the requirement that the thing loved be good is ambiguous between (i) the thing loved must be really good, (ii) the thing loved must appear to be good, and (iii) the thing loved must appear to be apparently good, and that my arguments are against (iii) but not (ii), and so fail to establish (i). But (iii) is not a possible reading of the requirement,



and my arguments against (ii) consist largely in drawing attention to Plato's explicit contrasts between (i) and (ii) in order to plump for (i).

15. McTighe ('Socrates on Desire', 212–15) argues that outside of the (he thinks) fallacious argument in the *Gorgias* discussed in Section 3.4, *boulesthai*, and in general verbs of desiring, relate the subject of desire to the apparent good, whereas the teleological lure exerted by the good is expressed in terms of 'for the sake of which' (*hou heneka*) or the notion of 'pursuit' (*diòkein* at *Republic* 505d–e). But the *Symposium* also seems to use a verb of desiring to describe the 'teleological lure' of the good on us.
16. There's a close linguistic parallel in Aristophanes' speech in the *Symposium*: the soul of each (of the halved-by-Zeus human beings) wants something it can't express but divines and guesses at what it wants (*all' allo ti... boulomenê... ho ou dunatai eipein, alla manteuetai ho bouletai, kai ainittetai*) (192c7–d2).
17. As Carone, who sees psychological eudaemonism in this passage, reads it, that the good is pursued in every action implies either that every desire is for the good or that every action is motivated by reason's (good-seeking) executive decision; she opts for the latter ('*Akrasia* in the *Protagoras* and *Republic*', at 133). But this limits 'every soul' in the passage to every adult human soul, and ends up dividing reason, for example when Leontius simultaneously looks at corpses and curses his desire to look (as argued in Section 4.2).
18. In the *Laches*, Nicias recalls that he has often heard Socrates saying that each of us is good in those things about which he is wise and bad in those things about which he is foolish (194d1–2), to which Socrates replies, 'But by god, you're speaking the truth, Nicias!' Thereafter (as elsewhere, with other virtues) Socrates goes on to drive the thesis that courage is some kind of knowledge into the ground, showing that the wisdom in question is either the wisdom of experts in some particular field (an implausible result), or, if it is knowledge of future good and evil, implies knowledge of good and evil generally, and so must be the whole of, rather than part of, virtue. Irwin (*Plato's Moral Theory*, 72) says that Socrates 'abruptly and without defense converts the search for a virtue into the search for a craft', thereby implying that he thinks knowledge is both necessary and sufficient for virtue—by contrast with virtue's involving the good condition of some non-rational motivations as well. However, Socrates' conversions of 'the search for a virtue into a search for craft' follow on his interlocutors' identification of the virtue with knowledge so that what Socrates himself contributes is the craft model of knowledge, not the identification of the virtue with knowledge. For example, at *Euthyphro* 14c–e Socrates is examining Euthyphro's definition of piety as knowledge of how to say and do what is pleasing to the gods,

14b; and at *Charmides* 165c–e, Charmides’ definition of temperance as self-knowledge, 164d—cf. ‘we are not investigating what I think but rather what you say’. And at *Gorgias* 460a–e, when Socrates asks whether the person who has learned just things is a just person, who does, and wants to do, [only] just things, he is asking whether Gorgias, who has just claimed that he teaches the student of oratory the good and bad, just and unjust, admirable and shameful, believes that justice is like carpentry and music, where the F person is the one who knows how to F. In these passages Socrates seems to use the craft analogy to put pressure on the identification of virtue and knowledge, in order to raise the question, ‘knowledge of what?’, rather than to express his own view that virtue has nothing to do with non-rational motivations.



# Bibliography

## Primary Sources (Editions, Translations, and Commentaries)

- Ackrill, J. L. (trans. and comm.), *Aristotle Categories and De Interpretatione* (Oxford, 1963).
- Adam, J. (ed. and comm.), *The Republic of Plato*, 2 vols (Cambridge, 1902).
- Allen, T. W. (ed.), *Homeri Opera*, vols III–IV (Oxford, 1917).
- Bandini, M. (ed.) and Dorion, L.-A. (trans.), *Memorables/Xenophon* (Paris, 2000).
- Barnes, J. (ed.), *The Complete Works of Aristotle* (Princeton, 1994).
- Barrett, W. S. (ed. and comm.), *Hippolytos* (Oxford, 1964).
- Basore, J. W. (trans.), *Seneca: Moral Essays*, vol. II (London, 1932).
- Buchheim, T. (ed. and comm.), *Reden, Fragmente und Testimonien/Gorgias von Leontini* (Hamburg, 1989).
- Burnet, J. (ed.), *Platonis Opera*, vols I–V (Oxford, 1900–7).
- Bywater, I. (ed.), *Aristotelis Ethica Nicomachea* (Oxford, 1894).
- Cooper, J. and Hutchinson, D. (eds), *Plato: Complete Works* (Indianapolis, 1997).
- Cornford, F. (trans. and comm.), *Plato's Cosmology* (New York, 1952).
- Diels, H. and Kranz, W. (eds), *Die Fragmente der Vorsokratiker* (Berlin, 1952).
- Dodds, E. (ed. and comm.), *Euripides Bacchae* (Oxford, 1960).
- Dodds, E. (ed. and comm.), *Gorgias* (Oxford, 1959).
- Faulkner, A. (ed. and comm.), *The Homeric Hymn to Aphrodite* (Oxford, 2008).
- Ferrari, G. R. F. (ed.) and Griffith, T. (trans.), *Plato: The Republic* (Cambridge, 2000).
- Fitzgerald, R. (trans.), *The Odyssey of Homer* (New York, 1963).
- Gagarin, M. (ed. and comm.), *Antiphon: The Speeches* (Cambridge, 1997).
- Gagarin, M. and Woodruff, P. (trans.), *Early Greek Political Thought from Homer to the Sophists* (Cambridge, 1995).
- Hackforth, R. (trans. and comm.), *Plato's Phaedrus* (Cambridge, 1952).
- Henry, P. and Schwyzler, H.-R. (eds), *Plotini Opera* (Oxford, 1964–82).
- Hicks, R. (trans.), *Diogenes Laertius Lives of Eminent Philosophers*, 2 vols (Cambridge, 1925).
- Irwin, T. (trans.), *Aristotle Nicomachean Ethics* (Indianapolis, 1985).
- Irwin, T. (trans. and comm.), *Plato: Gorgias* (Oxford, 1979).
- Jaeger, W. (ed.), *Aristotelis Metaphysica* (Oxford, 1957).
- Jouan, F. (ed. and trans.), *Euripide Tome VIII: Fragments 1<sup>re</sup> partie* (Paris, 1998).
- Lattimore, R. (trans.), *The Iliad of Homer* (Chicago, 1951).

- Lindskog, C. and Ziegler, K. (eds), *Plutarch: Vitae Parallelae* (Leipzig, 1957).
- Lloyd-Jones, H. and Wilson, N. G. (eds), *Sophoclis Fabulae* (Oxford, 1990).
- Lombardo, S. and Bell, K. (trans.), *Plato: Protagoras* (Indianapolis, 1992).
- Minio-Paluello, L. (ed.), *Aristotelis Categoriae et Liber De Interpretatione* (Oxford, 1949).
- Mirhady, D. and Too, Y. (trans.), *Isocrates I* (Austin, 2000).
- Monro, D. B. and Allen, T. W. (eds), *Homeri Opera*, vols I–II (Oxford, 1920).
- Murray, G. (ed.), *Euripidis Fabulae*, vols 2 and 3 (Oxford, 1902).
- Page, D. (ed.), *Aeschyli Septem Quae Supersunt Tragoedias* (Oxford, 1972).
- Page, D. L. (ed. and comm.), *Medea* (Oxford, 1938).
- Page, D. L. (ed.), *Poetae Melici Graeci* (Oxford, 1962).
- Pendrick, G. (ed., comm., trans.), *Antiphon the Sophist: The Fragments*. Cambridge Classical Texts and Commentaries 39 (Cambridge, 2002).
- Reeve, C. D. C. (trans.), *Plato: Republic* (Indianapolis, 2004).
- Shorey, P. (trans.), *The Republic/Plato*, 2 vols (Cambridge, MA, 1937).
- Taylor, C. C. W. (trans. and comm.), *Plato: Protagoras* (Oxford, 1976).
- Vlastos, G. (ed.) and Jowett, B. (trans.), *Plato: Protagoras* (New York, 1956).
- Yunis, H. (ed. and comm.), *Plato: Phaedrus* (Cambridge, 2011).

### Secondary Sources

- Allen, J., 'Dialectic and Virtue in Plato's *Protagoras*', in B. Reis (ed.), *The Virtuous Life in Greek Ethics* (Cambridge, 2006), 6–31.
- Allen, R. E. (ed.), *Studies in Plato's Metaphysics* (New York, 1965).
- Annas, J., *An Introduction to Plato's Republic* (Oxford, 1981).
- Annas, J., *Platonic Ethics Old and New* (Ithaca, 1999).
- Annas, J., 'Virtue and Law in the *Republic*', in R. Patterson, V. Karasmanis, and A. Hermann (eds), *Presocratics and Plato: Festschrift at Delphi in Honor of Charles Kahn* (Las Vegas, 2012), 165–82.
- Annas, J. and Rowe, C. (eds), *New Perspectives on Plato, Modern and Ancient* (Cambridge, MA, 2003).
- Anscombe, G. E. M., *Intention* (Ithaca, 1957).
- Asmis, E., "'Psychagogia' in Plato's 'Phaedrus'", *Illinois Classical Studies*, XI (1986), 153–72.
- Austin, J. L., 'A Plea for Excuses', *Proceedings of the Aristotelian Society*, ns. vol. 57 (1956), 1–30.
- Barney, R., 'Plato on the Desire for the Good', in S. Tenenbaum (ed.), *Desire, Practical Reason, and the Good* (Oxford, 2010), 34–64.
- Barney, R., Brennan, T., and Brittain, C. (eds), *Plato and the Divided Soul* (Cambridge, 2012).

- Bassi, J. and Euben, J. P. (eds), *When Worlds Elide: Classics, Politics, Culture* (Lanham, 2010).
- Benson, H., *Clitophon's Challenge: Dialectic in Plato's Meno, Phaedo, and Republic* (Oxford, 2015).
- Beresford, A., 'Nobody's Perfect: A New Text and Interpretation of Simonides PMG 542', *Classical Philology*, 103 (2008), 237–56.
- Betegh, G., 'Tale, Theology and Teleology in Plato's *Phaedo*', in C. Partenie (ed.), *Plato's Myths* (Cambridge, 2011), 77–100.
- Bett, R., 'Immortality and the Nature of the Soul in the *Phaedrus*', *Phronesis*, 31 (1986), 1–26.
- Blondell, R., *Plato and the Play of Character* (Cambridge, 2002).
- Bobonich, C., 'Akrasia and Agency in Plato's *Laws* and *Republic*', *Archiv für Geschichte der Philosophie*, 76 (1994), 3–36.
- Bobonich, C., 'Plato on Akrasia and Knowing Your Own Mind', in C. Bobonich and P. Destree (eds), *Akrasia in Greek Philosophy from Socrates to Plotinus* (Leiden, 2007), 41–60.
- Bobonich, C., *Plato's Utopia Recast: His Later Ethics and Politics* (Oxford, 2002).
- Bobonich, C. and Destree, P., *Akrasia in Greek Philosophy from Socrates to Plotinus* (Leiden, 2007).
- Bobzien, S., 'Choice and Moral Responsibility (NE iii.1–5)', in R. Polansky (ed.), *The Cambridge Companion to Aristotle's Nicomachean Ethics* (New York, 2014), 81–109.
- Bostock, D., *Plato's Phaedo* (Oxford, 1986).
- Brennan, T., 'The Spirited Part of the Soul and its Natural Object', in R. Barney, T. Brennan, and C. Brittain (eds), *Plato and the Divided Self* (Cambridge, 2012), 102–27.
- Brickhouse, T. and Smith, N., 'Socrates' Elenctic Psychology', *Synthese*, 92 (1992), 63–82.
- Brickhouse, T. and Smith, N., *Socratic Moral Psychology* (Cambridge, 2010).
- Broadie, S., *Nature and Divinity in Plato's Timaeus* (Cambridge, 2012).
- Brown, E., 'Justice and Compulsion for Plato's Philosopher-rulers', *Ancient Philosophy*, 20 (2000), 1–17.
- Brown, L., 'Glaucón's Challenge, Rational Egoism and Ordinary Morality', in D. Cairns, F.-G. Herrmann, and T. Penner (eds), *Pursuing the Good: Ethics and Metaphysics in Plato's Republic* (Edinburgh, 2008), 42–60.
- Brunschwig, J., Canto-Sperber, M., and Pellegrin, P. (eds), *Le Style de la Pensée: Recueils de textes en hommage à Jacques Brunschwig* (Paris, 2002).
- Bryan, J., 'Eikos in Plato's *Phaedrus*', in V. Wohl (ed.), *Probabilities, Hypotheticals and Counterfactuals in Ancient Greek Thought* (Cambridge, 2014), 30–46.
- Burnyeat, M. F., 'Culture and Society in Plato's *Republic*', *The Tanner Lectures on Human Values* 20 (Salt Lake City, 1999), 215–324.

- Burnyeat, M. F. (ed.), *Socratic Studies* (Cambridge, 1994).
- Burnyeat, M. F., 'The Truth of Tripartition', *Proceedings of the Aristotelian Society*, 106 (2006), 1–22.
- Calame, C. and Chartier, R. (eds), *Identités d'auteur dans l'Antiquité et la tradition européenne* (Grenoble, 2004).
- Callard, A., 'Everyone Desires the Good: Socrates' Protreptic Theory', *Review of Metaphysics*, 70 (2017), 617–44.
- Callard, A., 'Ignorance and Akrasia in the *Protagoras*', *Oxford Studies in Ancient Philosophy*, 47 (2014), 31–80.
- Calogero, G., 'Gorgias and the Socratic Principle *Nemo Sponte Peccat*', *Journal of Hellenic Studies*, 77 (1957), 12–17.
- Carone, G., 'Akrasia in the *Protagoras* and *Republic*: Did Plato Change His Mind?' *Oxford Studies in Ancient Philosophy*, 20 (2001), 107–28.
- Cherniss, H., 'The *Timaeus* in Relation to Plato's Later Dialogues', *American Journal of Philology*, 78 (1957), 225–66.
- Clark, J., 'The Strength of Knowledge in Plato's *Protagoras*', *Ancient Philosophy*, 32 (2012), 237–55.
- Cooper, J. M., 'Plato's Theory of Human Motivation', *History of Philosophy Quarterly*, 1 (1984), 3–21.
- Cooper, J. M., 'Socrates and Plato in Plato's *Gorgias*', in J. M. Cooper, *Reason and Emotion* (Princeton, 1999), 29–75.
- Davidson, D., 'Actions, Reasons, and Causes', in D. Davidson, *Essays on Actions and Events* (Oxford, 1980), 3–20.
- Davidson, D., *Essays on Actions and Events* (Oxford, 1980).
- Davidson, D., 'Mental Events', in D. Davidson, *Essays on Actions and Events* (Oxford, 1980), 207–24.
- Devereux, D., 'Socrates' Kantian Conception of Virtue', *Journal of the History of Philosophy*, 33 (1995), 381–408.
- Dray, W., *Laws and Explanation in History* (Oxford, 1957).
- Dyson, M., 'Knowledge and Hedonism in Plato's *Protagoras*', *Journal of Hellenic Studies*, 92 (1976), 32–45.
- Eliot, G., *Impressions of Theophrastus Such*, ed. N. Henry (Iowa City, 1994).
- Ferrari, G. R. F., *Listening to the Cicadas* (Cambridge, 1990).
- Fine, G., 'Knowledge and Belief in *Republic V–VII*', in G. Fine, *Plato on Knowledge and Forms* (Oxford, 2003), 85–116.
- Fine, G. (ed.), *The Oxford Handbook to Plato* (Oxford, 2008).
- Frede, M., 'Introduction', in S. Lombardo and K. Bell (trans.), *Plato: Protagoras* (Indiana, 1992), vii–xxxiii.
- Gallop, D., 'The Socratic Paradox in the *Protagoras*', *Phronesis*, 9 (1964), 117–29.
- Ganson, T., 'The Rational/Non-Rational Distinction in Plato's *Republic*', *Oxford Studies in Ancient Philosophy*, 36 (2009), 179–97.

- Geach, P. T., 'Plato's *Euthyphro*: Analysis and Commentary', *Monist*, 50 (1966), 369–82.
- Gill, C., 'The Body's Fault? Plato's *Timaeus* on Psychic Illness', in M. R. Wright (ed.), *Reason and Necessity: Essays on Plato's Timaeus* (London, 2000), 59–84.
- Gosling, J. C. B. and Taylor, C. C. W., *The Greeks on Pleasure* (Oxford, 1982).
- Gould, T., *The Ancient Quarrel Between Poetry and Philosophy* (Princeton, 1990).
- Gulley, N., 'The Interpretation of "No-One Does Wrong Willingly" in Plato's Dialogues', *Phronesis*, 10 (1965), 82–96.
- Gulley, N., *The Philosophy of Socrates* (New York, 1968).
- Gulley, N., 'Socrates' Thesis at *Protagoras* 358B–C', *Phoenix*, 25 (1971), 118–23.
- Harold, J., 'Infected by Evil', *Philosophical Explorations*, 8 (2005), 173–87.
- Hempel, C., *Aspects of Scientific Explanation* (Macmillan, 1965).
- Holmes, B., 'Body, Soul, and Medical Analogy in Plato', in K. Bassi and J. P. Euben, *When Worlds Elide: Classics, Politics, Culture* (Lanham, 2010), 345–85.
- Irwin, T. H., *Plato's Ethics* (Oxford, 1995).
- Irwin, T. H., *Plato's Moral Theory* (Oxford, 1977).
- Johansen, T. K., *Plato's Natural Philosophy: A Study of the Timaeus-Critias* (Cambridge, 2008).
- Kahn, C., 'On Platonic Chronology', in J. Annas and C. Rowe (eds), *New Perspectives on Plato, Modern and Ancient* (Cambridge, MA, 2003), 93–127.
- Kahn, C., 'Plato's Theory of Desire', *Review of Metaphysics*, 41 (1987), 77–103.
- Kahn, C., 'Proleptic Composition in the *Republic*, Or Why *Republic* I was never a Separate Dialogue', *Classical Quarterly*, 43 (1993), 131–42.
- Kahneman, D., Slovic, P., and Tversky, A. (eds), *Judgments Under Uncertainty: Heuristics and Biases* (Cambridge, 1982).
- Kamtekar, R., 'Ethics and Politics in Plato's Defense of Justice', in M. McPherran (ed.), *Plato's Republic: A Critical Guide* (Cambridge, 2010), 65–82.
- Kamtekar, R., 'Imperfect Virtue', *Ancient Philosophy*, 18 (1998), 315–39.
- Kamtekar, R., 'Plato on Education and Art', in G. Fine (ed.), *The Oxford Handbook to Plato* (Oxford, 2008), 336–59.
- Kamtekar, R., 'Plato on the Attribution of Conative Attitudes', *Archiv für Geschichte der Philosophie*, 88 (2006), 127–62.
- Kamtekar, R., 'Psychology and the Inculcation of Virtue in Plato's *Laws*', in C. Bobonich (ed.), *Plato's Laws: A Critical Guide* (Cambridge, 2013), 127–48.
- Kamtekar, R., 'Social Justice and Happiness in the *Republic*: Plato's Two Principles', *History of Political Thought*, 22 (2001), 189–220.
- Kamtekar, R., 'Speaking with the Same Voice as Reason: Personification in Plato's Psychology', *Oxford Studies in Ancient Philosophy*, 31 (2006), 167–202.
- Kamtekar, R., 'The Powers of Plato's Tripartite Psychology', *Proceedings of the Boston Area Colloquium in Ancient Philosophy*, 24 (2008), 127–50.



- Kamtekar, R., 'The Profession of Friendship: Callicles, Democratic Politics and Rhetorical Education in Plato's *Gorgias*', *Ancient Philosophy*, 25 (2005), 319–39.
- Kamtekar, R., 'The Soul's (After-)life', *Ancient Philosophy*, 36 (2016), 115–32.
- Kamtekar, R. (ed.), *Virtue and Happiness: Essays in Honour of Julia Annas, Oxford Studies in Ancient Philosophy Supplementary Volume* (Oxford, 2012).
- Knobe, J., 'Intentional Action in Folk Psychology: An Experimental Investigation', *Philosophical Psychology*, 16 (2003), 309–24.
- Kraut, R., 'Return to the Cave: Republic 519–21', *Proceedings of the Boston Area Colloquium in Ancient Philosophy*, 7 (1991), 43–62.
- Laks, A., "'Qu'importe qui parle." Sur l'anonymat platonicien et ses antécédents', in C. Calame and R. Chartier (eds), *Identités d'auteur dans l'Antiquité et la tradition européenne* (Grenoble, 2004), 99–117.
- LeBar, M. and Goldberg, N., 'Psychological Eudaemonism and Interpretation in Greek Ethics', in R. Kamtekar (ed.), *Virtue and Happiness: Essays in Honour of Julia Annas, Oxford Studies in Ancient Philosophy Supplementary Volume* (Oxford, 2012), 287–319.
- Lee, E. N., Mourelatos, A. P. D., and Rorty, R. M. (eds), *Exegesis and Argument: Essays in Honor of Gregory Vlastos* (Leiden, 1973).
- Liddell, H. G., Scott, R., and Jones, H. S., *Greek-English Lexicon*, 9th edn (Oxford, 1996).
- Lorenz, H., *The Brute Within: Appetitive Desire in Plato and Aristotle* (Oxford, 2006).
- MacKenzie, M. M., *Plato on Punishment* (Berkeley, 1981).
- McDaniel, K. and Bradley, B., 'Desires', *Mind*, 117 (2008), 267–302.
- McPherran, M. (ed.), *Plato's Republic: A Critical Guide* (Cambridge, 2010).
- McTighe, K., 'Socrates on Desire for the Good and the Unwillingness of Wrongdoing: *Gorgias* 466a–468e', *Phronesis*, 29 (1984), 193–236.
- Menn, S., 'On Plato's *Politeia*', *Proceedings of the Boston Area Colloquium in Ancient Philosophy*, 21 (2006), 1–55.
- Menn, S., 'Plato and the Method of Analysis', *Phronesis*, 47 (2002), 193–223.
- Menn, S., *Plato on God as Nous* (Carbondale, 1995).
- Meyer, S., *Aristotle on Moral Responsibility: Character and Cause* (Oxford, 2011).
- Moline, J., 'Plato on the Complexity of the Psyche', *Archiv für Geschichte der Philosophie*, 60 (1978), 1–26.
- Momigliano, A., *The Development of Greek Biography* (Cambridge, MA, 1993).
- Morris, M., 'Akrasia in the "Protagoras" and the "Republic"', *Phronesis*, 51 (2006), 195–229.
- Moss, J., 'Appearances and Calculations: Plato's Division of the Soul', *Oxford Studies in Ancient Philosophy*, 34 (2008), 36–68.
- Moss, J., 'Pleasure and Illusion in Plato', *Philosophy and Phenomenological Research*, 72 (2006), 503–35.

- Nails, D., 'Mouthpiece, schmouthpiece', in G. Press (ed.), *Who Speaks for Plato? Studies in Platonic Anonymity* (Lanham, 2000), 15–26.
- Netz, R., 'How Propositions Begin: Towards an Interpretation of *hupothesis* in Plato's Divided Line', *Hyperboreus*, 9 (2003), 295–317.
- Nussbaum, M., *Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy* (Cambridge, 1986).
- Obdrzalek, S., 'Contemplation and Self-Mastery in Plato's *Phaedrus*', *Oxford Studies in Ancient Philosophy*, 42 (2012), 77–107.
- O'Brien, D., *The Socratic Paradoxes and the Greek Mind* (Chapel Hill, 1967).
- Owen, G. E. L., 'The Place of the *Timaeus* in Plato's Dialogues', *Classical Quarterly*, 3 (1953), 79–95.
- Partenie, C. (ed.), *Plato's Myths* (Cambridge, 2011).
- Patterson, R., Karasmanis, V., and Hermann, A. (eds), *Presocratics and Plato: Festschrift at Delphi in Honor of Charles Kahn* (Las Vegas, 2012).
- Penner, T., 'Desire and Power in Socrates: The Argument of *Gorgias* 466a–468e', *Apeiron*, 24 (1991), 147–202.
- Penner, T., 'Knowledge vs. True Belief in the Socratic Psychology of Action', *Apeiron*, 29 (1996), 121–69.
- Penner, T., 'Plato and Davidson on the Parts of the Soul and Weakness of the Will', *Canadian Journal of Philosophy*, 16 (supplement) (1990), 35–74.
- Penner, T., 'Socrates on the Strength of Knowledge: *Protagoras* 351B–357E', *Archiv für Geschichte der Philosophie*, 79 (1997), 117–49.
- Penner, T. and Rowe, C., 'Desire for Good: Is the *Meno* Inconsistent with the *Gorgias*?' *Phronesis*, 39 (1994), 1–25.
- Polansky, R. (ed.), *The Cambridge Companion to Aristotle's Nicomachean Ethics* (Cambridge, 2014).
- Popper, K., *The Open Society and Its Enemies: The Spell of Plato* (Princeton, 1962).
- Press, G. (ed.), *Who Speaks for Plato? Studies in Platonic Anonymity* (Lanham, 2000).
- Price, A. W., *Mental Conflict* (New York, 1995).
- Reeve, C. D. C., *Philosopher-Kings: The Argument of Plato's Republic* (Princeton, 1988).
- Reis, B. (ed.), *The Virtuous Life in Greek Ethics* (Cambridge, 2006).
- Remes, P. and Sihvola, J. (eds), *Ancient Philosophy of the Self* (Dordrecht, 2008).
- Rickert, G., *HEKÔN and AKÔN in Early Greek Thought*, *American Classical Studies* 20 (Atlanta, 1989).
- Roberts, J., 'Plato on the Causes of Wrongdoing in the *Laws*', *Ancient Philosophy*, 7 (1987), 23–37.
- Robinson, R., *Plato's Earlier Dialectic* (Ithaca, 1941).
- Rorty, R., 'The Historiography of Philosophy: Four Genres', in R. Rorty, J. Schneewind, and Q. Skinner (eds), *Philosophy in History* (Cambridge, 1995), 49–75.

- Rorty, R., Schneewind, J., and Skinner, Q. (eds), *Philosophy in History* (Cambridge, 1995).
- Rudebusch, G., *Socrates, Pleasure, and Value* (Oxford, 1999).
- Sansone, D., 'Plato and Euripides', *Illinois Classical Studies*, 21 (1996), 35–67.
- Santas, G., 'Plato's *Protagoras* and Explanations of Weakness', *Philosophical Review*, 75 (1966), 3–33.
- Santas, G., 'The Socratic Paradoxes', *Philosophical Review*, 73 (1964), 147–64.
- Saunders, T., 'The Socratic Paradox in Plato's *Laws*', *Hermes*, 96 (1968), 421–34.
- Schleirmacher, F., *Philosophy of Plato and Aristotle*, trans. W. Dobson (New York, 1973).
- Schofield, M., 'Injury, Injustice and the Involuntary in the *Laws*', in R. Kamtekar (ed.), *Virtue and Happiness: Essays in Honour of Julia Annas*, *Oxford Studies in Ancient Philosophy Supplementary Volume* (Oxford, 2012), 103–14.
- Scott, D., *Plato: Meno* (Cambridge, 2006).
- Scott, D., 'Platonic Pessimism and Moral Education', *Oxford Studies in Ancient Philosophy*, 17 (1999), 15–36.
- Sedley, D., 'Aristotelian Relativities', in J. Brunschwig, M. Canto-Sperber, and P. Pellegrin (eds), *Le Style de la Pensée: Recueils de textes en hommage à Jacques Brunschwig* (Paris, 2002), 324–52.
- Sedley, D., *Creationism and its Critics in Antiquity* (Berkeley, 2007).
- Sedley, D., 'Platonic Causes', *Phronesis*, 43 (1998), 114–32.
- Sedley, D., *The Midwife of Platonism: Text and Subtext in Plato's Theaetetus* (Oxford, 2004).
- Segvic, H., 'No-One Errs Willingly: The Meaning of Socratic Intellectualism', *Oxford Studies in Ancient Philosophy*, 19 (2000), 1–45.
- Shields, C., 'Plato's Divided Soul', in M. McPherran (ed.), *A Critical Guide to Plato's Republic* (Cambridge, 2010), 147–70.
- Singpurwalla, R., 'Reasoning with the Irrational: Moral Psychology in the *Protagoras*', *Ancient Philosophy*, 26 (2006), 243–58.
- Singpurwalla, R., 'Why Spirit is the Natural Ally of Reason', *Oxford Studies in Ancient Philosophy*, 44 (2013), 41–65.
- Smyth, H. W., *Greek Grammar* (Cambridge, MA, 1956).
- Stalley, R., 'Plato's Argument for the Division of the Soul', *Phronesis*, 20 (1975), 110–28.
- Stocks, J. L., 'The Argument of Plato, *Protagoras* 351b–356c', *Classical Quarterly*, 7 (1913), 100–4.
- Tenenbaum, S. (ed.), *Desire, Practical Reason, and the Good* (Oxford, 2010).
- Tversky, A. and Kahneman, D., 'Judgments of and by Representativeness', in D. Kahneman, P. Slovic, and A. Tversky (eds), *Judgments Under Uncertainty: Heuristics and Biases* (Cambridge, 1982), 84–98.
- Vanderwaerdt, P., *The Socratic Movement* (Ithaca, 1994).

- Vasiliou, I., *Aiming at Virtue in Plato* (Cambridge, 2008).
- Vlastos, G., 'Introduction', in G. Vlastos (ed.) and B. Jowett (trans.), *Plato: Protagoras* (New York, 1956).
- Vlastos, G., 'Reasons and Causes in the *Phaedo*', *Philosophical Review*, 78 (1969), 291–325.
- Vlastos, G., 'Socrates on Akrasia', *Phoenix*, 23 (1969), 71–88.
- Vlastos, G., 'The Socratic Elenchus: Method is All', in M. Burnyeat (ed.), *Socratic Studies* (Cambridge, 1994), 1–37.
- Vlastos, G., 'The Theory of Social Justice in the *Polis* in Plato's *Republic*', in H. North (ed.), *Interpretations of Plato: A Swarthmore Symposium* (Leiden, 1977).
- Weiss, R., *Philosophers in the Republic: Plato's Two Paradigms* (Ithaca, 2012).
- Weiss, R., *The Socratic Paradox and Its Enemies* (Chicago, 2006).
- White, N., *Individual and Conflict in Greek Ethics* (Oxford, 2002).
- Whiting, J., 'Psychic Contingency in the Republic', in R. Barney, T. Brennan, and C. Brittain (eds), *Plato and the Divided Soul* (Cambridge, 2012), 174–208.
- Wilberding, J., 'Plato's Two Forms of Second-Best Morality', *Philosophical Review*, 118 (2009), 351–74.
- Wilburn, J., 'Akrasia and the Rule of Appetite in Plato's *Protagoras* and *Republic*', *Journal of Ancient Philosophy*, 8 (2014), 57–91.
- Wilburn, J., 'Courage and the Spirited Part of the Soul in Plato's *Republic*', *Philosopher's Imprint*, 15 (2015), 1–21.
- Wilburn, J., 'The Spirited Part of the Soul in Plato's *Timaeus*', *Journal of the History of Philosophy*, 52 (2014), 627–52.
- Williams, B. A. O., 'The Analogy of City and Soul in Plato's *Republic*', in E. N. Lee, A. P. D. Mourelatos, and R. Rorty (eds), *Exegesis and Argument: Essays in Honor of Gregory Vlastos* (Leiden, 1973), 196–206.
- Wohl, V. (ed.), *Probabilities, Hypotheticals and Counterfactuals in Ancient Greek Thought* (Cambridge, 2014).
- Wolfsdorf, D., 'Gorgias 466a4–468e5: Rhetoric's Inadequate Means', *Classical Philology*, 103 (2008), 103–34.
- Wolfsdorf, D., 'The Ridiculousness of Being Overcome by Pleasure: *Protagoras* 352b1–358d4', *Oxford Studies in Ancient Philosophy*, 31 (2006), 113–36.
- Woolf, R., 'Socratic Authority', in P. Remes and J. Sihvola (eds), *Ancient Philosophy of the Self* (Dordrecht, 2008), 77–107.
- Wright, M. R. (ed.), *Reason and Necessity: Essays on Plato's Timaeus* (London, 2000).



# Author Index

- Adam, J. 19 n. 28, 131 n. 1, 141 n. 19  
 Allen, J. 53 n. 35, 56, 56 n. 42, 57  
 Annas, J. 23 n. 35, 25 n. 39, 147, 147  
   n. 27, 154 n. 36  
 Anscombe, G. 91, 91 n. 44, 199 n. 13  
 Asmis, E. 177 n. 18  
 Austin, J. 72 n. 2
- Barney, R. 187 n. 3  
 Bell, K. 40 n. 8  
 Benson, H. 16 n. 24, 19 n. 27  
 Beresford, A. 73 n. 8  
 Betegh, G. 190 n. 6  
 Bett, R. 178 n. 20  
 Blondell, R. 7 n. 6, 13 n. 15, 13 n. 17  
 Bobonich, C. 45 n. 24, 144 n. 22, 146  
   n. 24, 146 n. 26, 147 n. 28, 147 n. 29,  
   148 n. 30, 176 n. 14  
 Bobzien, S. 112, 112 n. 74  
 Bostock, D. 192 n. 10  
 Bradley, B. 85 n. 31  
 Brennan, T. 174, 174 n. 11  
 Brickhouse, T. 2 n. 4, 54, 54 n. 38,  
   54 n. 40, 54 n. 41, 94 n. 49, 97 n. 54,  
   193 n. 11  
 Broadie, S. 196 n. 12  
 Brown, E. 26 n. 43  
 Brown, L. 44 n. 21  
 Bryan, J. 179 n. 21  
 Burnyeat, M. 17 n. 25, 132 n. 4, 165  
   n. 2, 172
- Callard, A. 47 n. 27, 92 n. 46, 100,  
   100 n. 60  
 Calogero, G. 44 n. 22, 73 n. 5  
 Carone, G. 2 n. 3, 134 n. 8, 140 n. 16, 141  
   n. 17, 152 n. 33, 202 n. 17  
 Cherniss, H. 7 n. 4  
 Clark, J. 43 n. 18, 48 n. 30  
 Cooper, J. 81 n. 20, 165 n. 2  
 Cornford, F. 107 n. 68
- de Havren, V. 35 n. 1  
 Devereux, D. 2 n. 4, 54 n. 38, 99 n. 57
- Diels, H. 76 n. 12  
 Diogenes Laertius 79 n. 16  
 Dodds, E. 89 n. 38, 90, 90 n. 43  
 Dray, W. 186–7, 187 n. 1, 187 n. 2, 190  
 Dyson, M. 43 n. 18
- Eliot, G. 9 n. 9
- Ferrari, G. 180 n. 24  
 Fine, G. 19 n. 27  
 Frede, M. 1 n. 1, 43 n. 18
- Gallop, D. 42 n. 14  
 Ganson, T. 143 n. 21  
 Geach, P. 15 n. 20  
 Gill, C. 107 n. 69  
 Goldberg, N. 187 n. 3  
 Gosling, J. 39 n. 7  
 Gould, T. 188 n. 5  
 Griffith, T. 23 n. 33  
 Grube, G. 23 n. 33, 133 n. 5  
 Gulley, N. 53 n. 35, 89 n. 39, 105 n. 65
- Hackforth, R. 178 n. 19  
 Harold, J. 188 n. 4  
 Hempel, C. 187 n. 1  
 Holmes, B. 97 n. 55
- Irwin, T. 1 n. 1, 2 n. 2, 13 n. 13,  
   13 n. 14, 13 n. 17, 15 n. 21, 26 n. 42,  
   42 n. 15, 44 n. 20, 59 n. 44, 81 n. 20,  
   83 n. 24, 83 n. 26, 85 n. 28, 134 n. 7,  
   137 n. 13, 137 n. 14, 139, 139 n. 15,  
   165 n. 3, 171 n. 5, 187 n. 3,  
   203 n. 18
- Johansen, T. 173 n. 7
- Kahn, C. 7 n. 3, 14 n. 19, 134 n. 7  
 Kahneman, D. 154 n. 37  
 Kamtekar, R. 23 n. 35, 36 n. 4, 85 n. 30,  
   90 n. 41, 95 n. 50, 110 n. 70, 151 n. 31,  
   152 n. 34, 153 n. 35, 156 n. 38,  
   156 n. 40, 187 n. 3

- Knobe, J. 72 n. 3  
 Kraut, R. 26 n. 41
- Laks, A. 7 n. 5  
 LeBar, M. 187 n. 3  
 Lombardo, S. 40 n. 8  
 Lorenz, H. 136 n. 12, 141 n. 19,  
     146, 146 n. 25, 175 n. 13, 176 n. 14
- MacKenzie, M. 107 n. 69  
 McDaniel, K. 85 n. 31  
 McTighe, K. 83 n. 24, 87 n. 33, 88 n. 34,  
     201 n. 15  
 Menn, S. 11 n. 11, 16 n. 23, 191 n. 8  
 Moline, J. 144 n. 22, 146 n. 24, 150–2  
 Momigliano, A. 6 n. 1  
 Morris, M. 43 n. 18  
 Moss, J. 97 n. 53, 143 n. 20, 146 n. 25
- Nails, D. 7 n. 6  
 Netz, R. 20 n. 30  
 North, H. 23 n. 35  
 Nussbaum, M. 43 n. 19, 177 n. 17,  
     179 n. 22, 180 n. 25
- Obdrzalek, S. 180 n. 23  
 O'Brien, D. 85 n. 29  
 Owen, G. 7 n. 4
- Penner, T. 2 n. 2, 44 n. 21, 52 n. 34,  
     85 n. 30, 100 n. 59, 187 n. 3  
 Popper, K. 23 n. 34  
 Price, A. W. 144 n. 23
- Reeve, C. 2 n. 2, 13 n. 13, 13 n. 14,  
     13 n. 16, 14 n. 18, 15 n. 21,  
     23 n. 33, 26 n. 40, 133 n. 5, 144 n. 22  
 Rickert, G. 81 n. 19  
 Robinson, R. 20 n. 29  
 Rorty, R. 8 n. 7  
 Rowe, C. 100 n. 59, 187 n. 3  
 Rudebusch, G. 39 n. 7
- Sansone, D. 79 n. 16  
 Santas, G. 43, 43 n. 16, 71 n. 1, 73 n. 6,  
     93 n. 48, 99 n. 57, 99 n. 58, 100 n. 61  
 Saunders, T. 110 n. 71  
 Sauvé Meyer, S. 112, 112 n. 73  
 Schleiermacher, F. 7 n. 4  
 Schofield, M. 109, 110 n. 70, 110 n. 72  
 Scott, D. 13 n. 15, 17 n. 25  
 Sedley, D. 8 n. 8, 133 n. 6, 156 n. 39,  
     191 n. 7  
 Segvic, H. 2 n. 4, 89 n. 40  
 Shields, C. 131 n. 2  
 Shorey, P. 23 n. 33  
 Singpurwalla, R. 51 n. 32, 54 n. 39,  
     165 n. 3  
 Smith, N. 2 n. 4, 54, 54 n. 38, 54 n. 40, 54  
     n. 41, 94 n. 49, 97 n. 54, 193 n. 11  
 Stalley, R. 131 n. 1  
 Stocks, J. 40 n. 10
- Taylor, C. 39 n. 7, 46 n. 26, 47 n. 28,  
     48 n. 29  
 Tversky, A. 154 n. 37
- Vasiliou, I. 15 n. 21, 15 n. 22  
 Vlastos, G. 17 n. 25, 23 n. 35, 36 n. 3,  
     41 n. 12, 42 n. 15, 53, 53 n. 36, 94,  
     94 n. 49, 191 n. 9
- Weiss, R. 8 n. 8, 42 n. 13, 79 n. 15  
 White, N. 44 n. 21  
 Whiting, J. 131 n. 2  
 Wilburn, J. 2 n. 3, 152 n. 33, 170 n. 4,  
     176 n. 15  
 Williams, B. 23 n. 34, 165 n. 1  
 Wolfsdorf, D. 47 n. 27, 47 n. 28, 48 n. 30,  
     85 n. 29, 88 n. 34  
 Woolf, R. 12 n. 12
- Yunis, H. 177 n. 18  
 Zeyl, D. 106 n. 67

# Index of Passages

## Aeschylus

### *Prometheus*

266 74

360 74

1364 74

## Antiphon

*Truth* 78, 80

7b 132 n. 3

## Aristotle

### *Categories*

8a31–32 133 n. 6

### *Metaphysics*

987b1–2 190 n. 6

1005b20 131 n. 1

1025a1–13 102 n. 63

1025a12–13 103

### *Nicomachean Ethics*

III.1 111, 113 n. 75

III.2–4 111, 113

III.5 70, 111

V.7–8 113

1105a17–21 113

1105a28–33 113

1105b5–10 113

1110a21 40 n. 10

1111b7–10 90

1111b20–25 90

1111b26 90

1111b26–30 90

1113a11 90

1113a15–22 90

1113b2–6 111

1113b6–7 111

1113b8–11 112

1113b10–14 112

1113b14–17 114

1113b17–21 114

1113b23–25 114

1113b30–33 114

1114a3–10 114

1114a11–12 115

1114a12–21 115

1114a15–31 115

1114a31–33 115

1114b1–16 115

1114b16–21 115

1114b21–25 115

1129b12–19 25

1135a20–23 113

1135a24–b2 113

1135b19–1136a3 114

1149a25–30 169

## *Politics*

II. 1264b23 7

## *Rhetoric*

### *Topics*

2.3 111a 83 n. 25

6.3 140b27 83 n. 25

6.8 146b9–12 83 n. 25

## Gorgias

*Defense of Palamedes* 73 n. 5

6–12 73 n. 5

13 73 n. 5

18 73 n. 5

19 44

25 44, 73 n. 5

*Encomium of Helen* 79, 81, 94

6 45 n. 24

1–6 75, 75 n. 11

8–13 106 n. 66

8–15 75, 76

15–19 75

19 77

## Euclid

*Optics* 49

## Euripides

### *Augē*

265.1 73 n. 7

### *Bacchae*

133–42 79



Euripides (*cont.*)

1173–1215 79

1251–79 79

*Electra*

1065–8 81 n. 19

*Heracles*

900–80 79

*Hippolytus*

373–83 45

1433 79

*Medea*

1042 44

1076–80 44

Herodotus

III.80.3 81 n. 22

Hippocrates*Peri Kardiēs*

Section 1 173 n. 9

Section 5 173 n. 9

Homer*Iliad*

45 n. 24

1.5 90

1.117 83 n. 25

1.258 90

2.202 90

23.594 83 n. 25

24.652 90

*Odyssey*

3.232 90

11.489 83 n. 25

12.350 83 n. 25

20.17–18 138

Isocrates*Against the Sophists*

15.11 180

14 180

18 180

*Antidosis*

15.266–278

15.271 180–1

*Helen*

5 181 n. 26

Plato*Alcibiades*

128e–129a 1

*Apology* 110

17b 101 n. 62

21c–22a 54 n. 38

24a 54 n. 38

23c 9 n. 9

25c 71

25c–26a 79 n. 15

25c–26a 89, 98

25d 71, 72

25e 44, 72

26a 72, 72 n. 4

29e–30e 54 n. 38

37a–b 79 n. 15

*Clitophon* 14 n. 19*Charmides*

164d 203 n. 18

167e 2, 89

167d–e 133

*Cratylus*

419d 173 n. 8

420b–c 95

*Critias* 7 n. 4*Crito*

47e–48a 1

49a 79 n. 15

*Letters*

6 n. 2

*Euthydemus*

273d 59

275b–277c 59

277d 59

277d–e 59

278c 59

278d 50

278d–282d 50

278e3–6 58

278e–282d 59

280a–b 59

282a 59

288d–292e 59

*Euthyphro* 13

6d–e 14

14c–e 203 n. 18

*Gorgias* 3, 70, 78, 89 n. 38, 93,

104, 114, 129, 135, 200,

201 n. 15

448e 14

452d-e 81  
456b 81  
460a-e 203 n. 18  
460b-c 81 n. 20  
461b-c 94  
463a-465e 97  
464a 97  
464b-465e 81  
464d 97  
465c-d 105  
466a 94  
466a10-b5 81  
466b 81  
466b3 81  
466b6-8 81  
466b9-e2 81  
466b9-10 81  
466d-e 93, 105  
466d7-8 81  
466e1 81  
466e6-7 81  
466e9-11 81  
466e13-467a1 81  
467a7 81  
467a10 81  
467c-d 83  
467d6-e1 83  
467e-468a 84  
468a-b 84  
468b1-4 84  
468c2-7 84  
468d1-5 91  
468d5-7 91  
468d7-e3 91  
468e1 81  
468e1-2 93  
468e3-5 91  
469e1 84  
470a 93  
470d 91  
471e-472c 95  
473d-474b 85  
474b 92, 93  
474d-475d 92  
475e-476a 95  
476b-479d 92  
480a 87  
480e-481a 92 n. 46  
481c 89  
481c5-d5 190  
481d-482c 95 n. 50

482b-c 95  
483a-e 78  
483c 91 n. 45  
491c 91 n. 45  
491d-e 85 n. 30  
491d-494a 96  
493a-b 54 n. 38  
493a-494a 145  
493d-494a 85 n. 30  
499d 97  
499e 87  
500a-501c 97  
500a1 87  
503c 97  
504d 82 n. 23  
507b 54 n. 38  
509d-e 101  
509e 1, 82  
509e-510a 101  
511c-512b 105

*Hippias Major*

296e 84 n. 27

*Hippias Minor* 11, 70, 75 n. 10

366c  
370d-e 75  
371d-e 75  
372a 75  
365d7 75  
366b 75  
366e-367a 101  
372d-e 101  
372e3-5 101  
373d-374b 102  
373a-376c 102  
373d2 102  
373d3-4 102  
373d5-7 102  
373e3 102  
373e4-5 102  
374b 104  
374d2-3 103  
374b-e 103  
374d2-3 103  
374e-375a 104  
375a 104  
375b 104  
375b-c 104  
375d 104  
375d-376c 101  
376a-b 101

Plato (*cont.*)

376c-d 101

376c1 101

*Laches*

192c-d 1

191d-e 54 n. 38

194d1-d 203 n. 18

*Laws*

2, 3, 7, 27, 112, 113, 114

644c-645a 145, 153 n. 35

644d 180 n. 24

644d-e 176

644e-645a 171

714c4 24

731c 107

860d 107

860d1-861a2 107

861c2-862d4 108

861c3 110

863e-864a 109

896e9-897a4 148

904c 115 n. 77

*Lysis* 13

208e 40 n. 10

*Menexenus* 7

237a 40 n. 10

*Meno* 16, 17, 17 n. 25, 19, 20, 21,

37, 135

70a 9

71b 14, 15

72a-b 14

72c 97

73e-74a 14

75c-d 10

77b4-5 97

77b7-c2 97

77c1-2 97

77d-e 100

77d7-e3 99

78a 97

78a4 99 n. 58

78a4-9 98

78a9 100

78b3 98

78b5 97, 98

78e 101

79d-e 14

86d 15

86e-89a 15

86e3-4 15

87b3-4 15

*Parmenides* 7 n. 4

128c-d 55

*Phaedo* 7 n. 4, 17, 20, 156 n. 41, 187

65b 135 n. 10

65c 135 n. 10

66a 135 n. 10

66b-c 193

66c 135 n. 10

67a 135 n. 10

69e 40 n. 10

80a 135 n. 10

83c-d 193

83d 135 n. 10

94c-e 135 n. 10, 145

96e-101d 133 n. 6

97a-b 42, 191

98c 192

98c-99b 191-2

98e-99a 42

99a 192

100a-c 194

100d 194

101a-b 42

101d 16

102d-e 195

105c 195

105d 178 n. 20

110a 16

*Phaedrus* 4, 6, 7 n. 4, 10, 27, 139, 151,

166, 172, 176, 177 n. 17, 179 n. 21

230a4-5 130

237b-241d 177

237d 177

237e 177

237e-238a 177

238a 10

238b7-c4 177

239a-240a 177

239b4 177 n. 18

241a 177

244a 178

245c-256e 178

245c5 178

245c7-8 178

245dc8-d1 178

245d1-4 178

245e2-4 178

245e4-6 178

- 246a 9, 10, 130, 179  
246a-b 179  
246c 179  
247a 179  
247a-b 179  
247d 17  
248a-b 179  
248d-e 179  
249a-b 179  
249d 178, 180  
251b7 179  
253d-e 145, 179  
253e 180  
253e-254a 179, 180  
254b-d 145, 180  
254c 180, 180 n. 24  
254c-d 145  
254e 180  
262a-b 10  
265b-c 92  
270d 131  
271d 9-10  
278e-279a 180  
*Philebus* 7 n. 4, 139  
20d8-9 201  
20d 88  
35a-c 157  
38e-39c 176  
48b 189, 195  
48b-50a 29 n. 9  
*Protagoras* 2, 3, 4, 6, 7 n. 4, 35-6, 51,  
54, 55, 69, 78, 97 n. 55, 129, 139,  
140, 143, 155, 186  
309a-328d 36  
319a 36  
319a4 36  
319b-320b 36 n. 4  
320d-323a 36  
322c-d 107 n. 68  
323c-328a 37  
324a6-b5 107 n. 68  
325d 38  
326c 38  
327c-e 107 n. 68  
328a-d 37  
328b 36  
328e 37  
329e-330b 37  
330c-334c 36  
331c 36  
331e 36  
332a-e 196  
332a-333a 41, 55  
332b 108  
333c 36  
334a-c 36 n. 4, 49 n. 31  
334d-349a 38  
339a-347a 73  
345d-e 73  
349b 38  
349b-351a 36  
349c-351a 38  
349e-350c 38  
351a 38  
351b-360d 36  
351b-e 39  
351c 39  
352b 41, 45 n. 23  
352d-e 41  
352d7 38  
353a1 41  
353b1-3 38  
353c4-356a5 39  
353d-e 39  
354b 39  
354e-355a 43  
355a5-356a5 41  
356a-b 49  
356a5-8 47  
356a5-357e8 39  
356b1-c1 48  
356c 49  
356c-357a 48  
356c5 49  
356d 54 n. 39  
356d3-e2 50  
357a-b 50  
357d-e 52  
357e 36, 52  
358a 50  
358a5-360d9 40  
358b8-c2 52  
358c 1  
358c-360d 39  
358c3-4 52  
358c4-5 52  
358c4-d4 52  
358d 56  
358d1-2 53  
358e 56  
359c-e 56

Plato (*cont.*)

- 359e-360a 56  
 360b 56  
 360c 56  
 360c 56  
 360c-d 1  
 360d 57  
  
*Republic* 2, 3, 6, 7 n. 4, 8, 13, 14, 15,  
 17, 19, 20, 21, 22, 26, 89, 115, 140,  
 150, 165, 175 n. 12, 176, 177 n.  
 17, 186-7, 193, 200, 202  
 330d 14 n. 19  
 331b 73 n. 7  
 334b-335a 199 n. 14  
 335c-d 42  
 338d 18, 24  
 339c 24  
 340a-c 88  
 340d-341a 24  
 340e-341d 14 n. 19  
 341c-342e 24  
 343b 24  
 346a-b 151  
 346c-d 24  
 346e-347d 14 n. 19  
 351e 151  
 354a13-c3 14  
 354a-c 12  
 356a-b 149  
 358b 151  
 358b7-9 78  
 358c1-4 78  
 359c1-2 78  
 359c3-6 78  
 358c7-8 44  
 358e-359b 18  
 358e-359c 131  
 359a 18, 23  
 359c4-6 131  
 360d 44, 106  
 367a 151  
 368e 17, 23  
 368a 23  
 369b 18, 23  
 369b-370a 18  
 369e-370a 22 n. 32, 23  
 371e 23 n. 36  
 372a-d 132  
 372e 22 n. 32, 23 n. 36  
 373a 132  
 375e-376b 168  
 376d 23 n. 36  
  
 379d 195  
 380b-c 25, 106  
 382b 53 n. 35, 105  
 383c 25  
 398d 170  
 399e 132  
 402d-403c 178  
 403b-c 25  
 409d-410a 25  
 410b-c 170  
 410b-411e 166  
 410d-e 170  
 411a-b 170  
 411b 145  
 412d-e 25  
 412e 106  
 413a 106  
 413b-d 106  
 413b-e 170  
 417b 25  
 429b-430b 151  
 430e6 147  
 431a 147  
 431a6 147  
 432d-433e 18  
 433b-c 18  
 433e-433a 18  
 435a-b 17, 21  
 435b-441c 17  
 435c-d 9, 137  
 436a 129, 141  
 436b-441c 129  
 436b8-c1 131  
 436c 131 n. 1  
 436d-e 131 n. 1  
 437b1-4 132  
 437b7-c6 133  
 437c 136  
 437d-e 133  
 438a 133-4  
 438b-e 133  
 439a-b 136  
 439c-d 136  
 439d 136  
 439e-440a 138, 166  
 440a-b 167  
 440b5-d3 167  
 440d 168  
 440e-441c 166  
 441a-b 138  
 441a2-3 167  
 441b 138, 168  
 441b-c 168

- 441e 170  
441e-442a 17  
442b-c 169  
442c 143, 146  
442c-d 144  
442d1 147  
443b 151  
443c-445b 17  
443d7 137  
444d 168  
449a 23  
458c-d 25  
471c 25  
477c 151  
477c-478b 151 n. 32  
485d 180 n. 24  
492d-493d 188  
504a-b 9  
504b 137  
505d-e 47 n. 27, 88, 139, 140, 143,  
155, 157, 165 n. 3, 201  
506a 151 n. 32  
506b-e 151 n. 32  
506c 10 n. 10, 151 n. 32  
506e-509b 10 n. 10  
508a-509b 151  
509c 151 n. 32  
509c3 10 n. 10  
509d-10a 89  
510b2-511a8 20  
510 21  
519d-520a 25 n. 38  
520b-c 26  
520c 151 n. 32  
521d 151  
523e-524a 133 n. 6  
532c 151  
533a 10 n. 10, 151  
537d-539a 169  
547b-c 174  
548b 174  
548d-550b 166  
548d-548a 9  
549a 171  
549c-550b 171  
553a-d 171, 194  
553a-c 194  
553c-d 171  
553d 152  
554b-e 137  
554c 175  
554c-d 148  
554c-d 145  
554d1 148  
558d-559c 136 n. 11  
559d-561d 171  
560b 152  
561b 152  
571a-573b 171  
572e 177  
580d-581c 130, 138, 144  
580d8 151  
580e-581a 145, 146  
581a-b 171  
581c-582e 154  
586d-e 151  
586e 17  
588b 151  
588b-589b 153  
595b-6 188  
600e5 188  
602a-b 188  
602c 141  
602c-604d 129  
602d 141, 146  
602e 141, 144  
602e4-6 141  
603a 142  
603e-604a 142  
604a 142  
604a-c 188  
604b 142  
604b-c 142, 188  
604b-d 144  
604d 142  
604d8-9 137  
605b-c 189  
605d 188  
606a8-b1 137  
606b 149, 188  
606b7-8 137  
606d 149  
611b-612a 145  
611c2 171  
611d8 171  
611e-612a 9  
617e 106  
*Sophist* 7 n. 4, 139  
228c-d 95 n. 51  
230b 131 n. 1  
*Statesman* 139  
*Symposium* 8  
192c7-d2 201 n. 16  
195a-196b 197  
200b4-5 199

Plato (*cont.*)

200c-d 199  
 201c-206a 197  
 204a6-7 199  
 204e2-205a3 198  
 205a-206a 198  
 205e-206a 199  
 206b 200  
 206b-209e 198  
 206e2-5 83 n. 25  
 207b-c 200  
 207c-d 200  
 210a-212b 198

*Theaetetus* 7 n. 4, 8, 36 n. 2

146d 14  
 163d-164d  
 165b-c 131 n. 1  
 166b 131 n. 1  
 166d-167a 36 n. 4  
 167a-d 88  
 171e-172b 88  
 177d-178a 88  
 184e-186d 176

*Timaeus* 2, 3, 6, 7 n. 4, 27, 108, 115,  
 143, 166, 172, 190, 193

29c 9  
 29e-30b 196  
 37a-b 155  
 39e-40a 197  
 41b-d 155  
 41d 155  
 41e-42d 155  
 42a-b 173 n. 6  
 42e-43a 155  
 43a-e 197  
 43d 155  
 46e-47e 156 n. 41  
 47a-c 173  
 69c-71e 155  
 69c9 173  
 69d5 173  
 69d7 173  
 70a 145  
 70a5 173  
 70a7 173  
 70b 173, 173 n. 9  
 70c 173 n. 9

71a-b 175  
 71a-d 180, 180 n. 24  
 71a5 174  
 71d-e 145  
 77b-c 145, 156, 157, 176  
 86b-87b 194  
 86b3-4 107  
 86c2-3 106  
 86d-e 107  
 86d7-e1 106  
 86e2-3 107  
 87a7 106  
 87b1-3 106  
 87b4 106  
 87c-89d 107  
 91d-92c 156  
 92a 16

Plotinus

*Enneads*

1.4.6.19 91

Plutarch

*Pericles*

36.3 80 n. 18

Seneca

*On Peace of Mind*

13.2-14.2 86

Simonides

fr. 541 PMG 74

Sophocles

*Oedipus at Colonus*

960-99 80

973-6 80

*Philoctetes*

116 101 n. 62

*Women of Trachis*

727-8 79

935 79

1123 79

Xenophon

*Memorabilia*

1.4 190 n. 6

# Subject Index

## action

akratic 2–4, 40–59, 47 n. 27, 139–43, 147, 169, 177

contrary to knowledge 2, 38, 41–2, 49–52, 56

vs. contrary to belief 2, 41, 53  
(*see also* conflict)

character and 107 n. 68, 107–15, 114 n. 76

explanation of (*see under explanation*)  
instrumental 70, 83–8, 90, 146

involuntary, *see* unwilling

unintentional, *see* unwilling

Adeimantus 9, 11 n. 11, 12, 13 n. 15, 17, 22, 23 n. 36, 106

Anaxagoras 190 n. 6, 191, 194, 197

appearance 47, 90, 165 n. 3, 188, 190, 199–201

vs. calculation or measurement 48–52, 141–3, 154, 188

## argument-form

hypothetical 12, 15–21, 35–8

protreptic 39–40, 48–52, 58–9

refutation (elenchus) 12–13, 35–6, 37, 39, 40–3, 45, 47, 50–1, 53–4, 57, 59, 94

Aristophanes 79 n. 16, 199, 201 n. 16

Aristotle 7, 11, 19, 25, 40 n. 10, 70,

72 n. 3, 81 n. 20, 83 n. 25,

90, 90 n. 42, 102 n. 63, 103,

111–15, 131 n. 1, 133 n. 6,

169, 190 n. 6

Athenian Visitor 6, 107–10

Callicles 9, 10, 73 n. 6, 78, 79 n. 16,

83, 87, 88, 88 n. 34, 91 n. 45,

95 n. 50, 96–7, 190

cause, *see* explanation

character (*state*) 70, 75, 77, 107,

109–11, 130, 147, 151–2,

171–2

relation to action 106 n. 68, 107–15, 114 n. 76

characters (*in dialogues*), characterization,  
*see under* dialogue form

## conflict

among virtues 37

psychic 4, 40–59, 129, 136–8, 140–3, 146, 166, 171 n. 5

correlatives 133, 136

## definition

applies to Form 17

priority of 14–15

hypothesis and 15

of soul 178–9

desires, *see under* motivation

## dialogues, dialogue form

characters, characterization 6, 6 n. 1, 7–10, 12, 14

chronology 2, 7, 7 n. 3–4, 17 n. 25, 190 n. 6

dialectical dependence 3, 10–12, 22, 25–6, 132, 138, 194; *see also under* hypothesis

vs. Dialectical 10 n. 10, 11, 13, 18

drama *and* 11–22, 144

hypothesis *and* 12, 21–2, 25–6

two-level readings 8

Eleatic Visitor 6

Empedocles 180

expertise (technè) 24–5, 36–8, 82, 88, 97 n. 55, 97, 101–5

Euthyphro 9

## explanation (*or* cause)

by character 195–7

by formal cause 194–7

by goodness 6, 52–4, 155–7, 187, 190, 196–7, 199 n. 13

by intelligence (nous) 187,

190–3, 191 n. 7, 192 n. 10,

196–7

by parts 139, 147

genetic 193–4

hypothetical (*see under* hypothesis)

psychological 5–6, 35, 55, 139, 197



- explanation (*or cause*) (*cont.*)  
   psychological eudaemonism 39–40,  
     42 n. 15, 43–4, 43 n. 18, 51–9,  
     69–71, 81 n. 20, 83, 83 n. 26,  
     129–30, 139–40, 149, 155,  
     186–7, 187 n. 3, 198, 201  
   vs. rational eudaemonism 44,  
     44 n. 20, 59 n. 44  
   teleological 4, 6, 130, 135–6, 155–7,  
     187, 190 n. 6, 194, 197, 201
- Glaucon 9–10, 11 n. 11, 12, 17–18, 20,  
   22–6, 44, 78, 79 n. 15, 106,  
   131–2, 138, 167–8
- good (*see also under* explanation)  
   desire for (*see also under* motivation)  
     2–4, 40–4, 48–55, 70, 104–5,  
     129–39, 143, 147–50, 154–7,  
     165 n. 3, 177, 186–7, 198, 202  
   in instrumental actions 81–90  
   real vs. apparent 87–9, 96–100  
   realism vs. conventionalism about 201–2
- hedonism  
   ethical 3, 39–40, 43, 43 n. 19,  
     53–4, 186  
   psychological 42–4, 43 n. 18, 47, 54,  
     57, 149
- hypothesis  
   vs. belief 12  
   method of 3, 15–21, 194  
   consequences and 19, 26  
   investigation on a 15–21, 35–6, 54  
   higher hypothesis 3, 16,  
     16 n. 24, 35  
   investigation of justice  
     and 17–22, 25  
   city-soul analogy *and* 17–22  
   *in geometry* 15–16, 16 n. 23, 19–20  
   about justice 17–22, 25–6  
   about virtue 15–17, 19, 21,  
     35–8, 186  
   *in thought* (*dianoia*) 10 n. 10,  
     19, 19 n. 27
- ignorance, *action due to, see under*  
   unwilling
- injustice (*see also* wrongdoing) 12–14,  
   17, 23, 44, 75, 78, 87–8, 91–5,  
   101–3, 165, 175–6, 189  
   vs. injury 109–14
- of character 75, 107–14  
     vs. unjust act, act of injustice 75,  
       108–9
- intensional contexts 40 n. 10, 46–7
- Isocrates 180–1
- knowledge 177 n. 18, 180–1, 187 n. 3,  
   188, 190 n. 6, 193, 196, 198,  
   200–1, 203, 203 n. 18  
   action contrary to, *see* action, akratic  
   vs. belief 43, 43 n. 18, 47 n. 27,  
     52 n. 34, 53  
   factive 45–6, 53  
   priority of definition *and* 15–17  
   virtue *and* 1, 3, 6, 15, 16 n. 23, 19–21,  
     35–9, 38 n. 5, 53, 69, 203
- Meno 9, 14–15
- motivation  
   anger 44–5, 45, 109, 114, 137–8,  
     140–1, 145, 153, 155, 165–9,  
     171, 173–6  
   force model 3–4, 35, 38, 40–1, 44–8,  
     57–8, 129–33, 134 n. 9, 136–8,  
     140–3, 147–8, 153  
   good-directed vs. good-indifferent 1–2,  
     96, 99, 129, 131–40, 143–57  
   multiple sources (*see also* soul-parts)  
     13, 130–2, 136–40,  
     142–3, 165  
   rational vs. non-rational 2, 2 n. 4,  
     4, 13, 22, 54, 129, 139–40,  
     143–5, 149–52, 165, 176,  
     188–9, 203  
   psychological vs. kinetic  
     characterization of 136,  
     148–52  
   desire 165 n. 3, 166, 169, 171,  
     173 n. 6, 173, 176–7, 180,  
     186–7, 187 n. 3, 189–90, 193,  
     197–200, 199 n. 13, 201 n. 15,  
     202  
   appetite (*epithumia*,  
     *epithumein*) 83, 96–8, 101,  
     130–2, 167, 172–8, 175 n.  
     13–15 (*see also* soul-parts,  
     appetitive)  
   attribution of 4, 86, 93–6, 129,  
     133–5, 138, 199–200  
   compared to attribution of  
     belief 94–6, 135 n. 10

- cancelled 85, 140, 200  
 conditional 85–9, 105–11, 129  
 for good vs. apparent good 48–50,  
   87–90, 96–100, 129, 201  
 instrumental 70, 83–90, 89 n. 37,  
   93, 135  
 rational vs. non-rational 1, 13, 54,  
   132, 143–54  
 wish (*boulesthai*) 2, 58, 83 n. 25, 87  
   n. 33, 89–91, 98, 132, 136, 147,  
   198, 201 n. 15
- Parmenides 6, 9, 55, 180  
 Pericles 80 n. 18  
 poetry 13, 25, 38 n. 5, 76, 170,  
   187–90, 200  
 Polus 14, 81–4, 86–8, 91–5  
 power 11, 19, 25, 37, 73 n. 5, 75–8, 81–6,  
   91 n. 45, 93–6, 177  
   conceptions of 81 n. 20, 89, 91,  
   101–5  
 powers of the soul 131, 137, 151–2, 174,  
   175, 175 n. 13, 176, 179–81  
 Proclus 19 n. 28  
 Protagoras 35–40, 50–4, 80 n. 18
- refutation (*see under* argument-form)
- soul-parts 13, 129, 165–77, 171 n. 5, 173  
   n. 7, 174 n. 11, 175 n. 12–14,  
   177 n. 17–18, 179–81, 188–90,  
   193–4, 202  
 agent-like 4, 130, 144–7, 149, 152–3,  
   186, 189  
 appetitive 136–9, 141, 143–6, 150–4,  
   165, 167, 172–6, 190  
 arguments for 129, 131, 136–8, 140–2  
   conflict *and* 129, 136–8, 141–2,  
   146, 166  
   principle of opposites *and* 131–2,  
   136–7, 166, 168, 190  
   correlatives *and* 131  
 good-directed 4, 130, 140, 143–4,  
   147–57, 165, 165 n. 3,  
   177 n. 17, 181, 186, 189, 202  
 spirited  
   honour-loving 137–8, 144–5,  
   150–4, 165–6, 169, 171, 174–5,  
   179, 181  
   vs. reason-responsive 165–72, 180  
 personified 144, 152–4, 175, 189
- reasoning 136, 142–3, 146, 165–6, 171  
   n. 5, 177 n. 17, 180, 188, 202
- teleology (*see under* explanation,  
   teleological, by nous, by  
   goodness)
- Theaetetus 10, 14  
 Thrasy machus 9, 11–14, 18, 22–5,  
   44, 88  
 Timaeus 6, 106–7, 115, 155–7
- unwilling
- action 58, 72 n. 2, 72 n. 3, 73–4,  
     81 n. 19, 89–92, 101–5  
   attributability to agents 42, 80–1  
   contrary to nature 4, 53, 78–9, 157  
   due to compulsion 40–1, 44–7,  
     74–81  
   due to ignorance 41, 48–54, 69–73,  
     93, 101, 105–11, 114–15  
   ignorance enables  
     compulsion 79–81  
   ignorance *and* action-  
     description 80–1  
   wanting or wishing *and* 40–2, 82,  
     92–3, 105–11 (*see also* wish  
     *under* motivation)  
   *see also under* wrongdoing  
   condition 96 n. 52, 105–7, 108  
   vs. willing in the law 107–11
- vice (*see also under* unwilling condition)
- 1, 3, 14, 53, 71–2, 97–9,  
   100 n. 59, 103, 105, 107,  
   107 n. 69, 111–15, 144,  
   152, 172
- virtue 1, 12–15, 17–19, 35–41, 50–3,  
   55–7, 59, 72, 86, 97–8, 101–15,  
   171–2, 174, 174 n. 10, 186,  
   188–9, 191 n. 7, 194–5, 197,  
   203, 203 n. 18
- unity of 37  
   identical to wisdom 38
- wrongdoing (*see also under* injustice *and*  
   unwilling) 1–5, 52–4, 69–74,  
   79 n. 15, 92–3, 186  
   responsibility for in Aristotle 111–15  
   willing 101–5
- Zeno 55