

Plato's *Republic*

An Introduction

SEAN McALEER





<https://www.openbookpublishers.com>

© 2020 Sean McAleer



This work is licensed under a Creative Commons Attribution 4.0 International license (CC BY 4.0). This license allows you to share, copy, distribute and transmit the text; to adapt the text and to make commercial use of the text providing attribution is made to the authors (but not in any way that suggests that they endorse you or your use of the work). Attribution should include the following information:

Sean McAleer, *Plato's Republic: An Introduction* (Cambridge, UK: Open Book Publishers, 2020), <https://doi.org/10.11647/OBP.0229>

In order to access detailed and updated information on the license, please visit, <https://doi.org/10.11647/OBP.0229#copyright>

Further details about CC BY licenses are available at <https://creativecommons.org/licenses/by/4.0/>

All external links were active at the time of publication unless otherwise stated and have been archived via the Internet Archive Wayback Machine at <https://archive.org/web>

Updated digital material and resources associated with this volume are available at <https://doi.org/10.11647/OBP.0229#resources>

Every effort has been made to identify and contact copyright holders and any omission or error will be corrected if notification is made to the publisher.

ISBN Paperback: 9781800640535

ISBN Hardback: 9781800640542

ISBN Digital (PDF): 9781800640559

ISBN Digital ebook (epub): 9781800640566

ISBN Digital ebook (mobi): 9781800640573

ISBN XML: 9781800640580

DOI: 10.11647/OBP.0229

Cover image: Anselm Feuerbach, *The Symposium* (1873). Berlin State Museum, Wikimedia, [https://commons.wikimedia.org/wiki/File:Anselm_Feuerbach_-_Das_Gastmahl._Nach_Platon_\(zweite_Fassung\)_-_Google_Art_Project.jpg](https://commons.wikimedia.org/wiki/File:Anselm_Feuerbach_-_Das_Gastmahl._Nach_Platon_(zweite_Fassung)_-_Google_Art_Project.jpg). Public domain.

Cover design by Anna Gatti.

11. The Allegory of the Cave

Book VII

The Allegory of the Cave is arguably the most famous part of the *Republic*. Although it is clearly related to the Sun and Divided Line analogies (indeed, Socrates explicitly connects the Cave and the Sun at 7.517bc), Plato marks its special status by opening Book VII with it, emphasizing its importance typographically, so to speak (he will do much the same thing in Book IX with the discussion of the tyrannical soul). Although an allegory is sometimes defined as a symbolic narrative that can be interpreted as having a hidden meaning, Plato is not cagey about the Cave Allegory's meaning: it is about 'the effect of education (*παιδεία* [*paideia*]) and the lack of it on our nature' (7.514a). Given how visual the allegory is, many readers will find it helpful to draw themselves a diagram of it.

Education, the Allegory's topic, is not what most people think it is, says Plato: it is not 'putting knowledge into souls that lack it' (7.518b). Though education sometimes requires that kind of transmission of knowledge from teacher to student, this is not its essence, which instead is 'turning the whole soul' (7.518d)—turning it around, ultimately toward the Form of the good. *Education as turning around* is a powerful metaphor, capturing the way in which learning involves gaining new perspectives, seeing everyday things and events from new points of view. Everyone, Plato insists, is capable of education in this sense (7.518c). But not everyone is capable of making it out of the Cave into the intelligible world of the Forms, just as not everyone is capable of winning a Nobel Prize in Physics or an Olympic medal in Figure Skating. Nonetheless,

everyone has the capacity to be educated, to turn their soul from what is less real toward what is more real.

Stages in the Cave Allegory

I count six distinct stages in the Cave Allegory. While such divisions are always prey to arbitrariness and subjective preference, I hope that the division I offer sheds light on what Plato is up to here.

In the first stage, the cave's residents are prisoners, chained to their seats and unable to move not only their bodies but—crucially—their heads. They can only look straight ahead, and thus have only one perspective on what they see on the cave's wall. What they see are the shadows of a sort of puppet show taking place behind them, with shadows cast by the light of a fire. The puppets are various artifacts: 'statues of people and other animals, made out of stone, wood, and every material' (7.514b). The prisoners watch the shadow-play, ignorant of the true nature of what they see: they 'believe that the truth is nothing other than the shadows of those artifacts' (7.515c). They take for reality what is a mere image of it. Some readers will have already noticed that Stage One is parallel to the lowest section of the Divided Line (segment a), the objects of which are images and shadows.

In the second stage, one of the prisoners is freed from their bonds. Plato does not tell us by whom or how; we are left to wonder whether the prisoner was saved by human agency or by the natural decay of their fetters. There is reason to think it is the former, since the freed prisoner is 'suddenly compelled to stand up, turn his head, walk, and look up toward the light' (7.515c), and somebody else seems to be doing the compelling. This is not the only time Plato connects education with compulsion, with being forced to turn one's head and gain a new perspective. Nor is it the only time when the head-turning that constitutes education will be painful. When the freed prisoner is forced to look at the shadow-casting fire that until this moment they were unaware of, they will be 'pained and dazzled and unable to see the things whose shadows they had seen before' (7.515c). They will probably not like the experience at all, even though in being freed from their fetters they are thereby 'cured of [their] ignorance' (7.515c)—not merely freed but cured, as if ignorance is a disease. It is a comfortable disease, to borrow a

phrase from e. e. cummings, for it is a world the cave-dweller is familiar with and comfortable in. Turned around and out of their comfort zone, they are unable to recognize the shadow-casting puppets, despite their skill at recognizing the shadows the puppets cast on the wall. Although the artifacts are, like any sensible particulars, not fully real, they are *more* real than the shadows they cast. Thus in looking at the shadow-casting artifacts the freed prisoner is 'a bit closer to the things that are and is turned towards things that are more' (7.515d); this is the *existential* sense of the verb 'to be' that we distinguished earlier: the prisoner is closer to the things that are real—that exist—and indeed is coming closer to the things that are fully real: the Forms. While not everyone is capable of making it out of the Cave, Plato thinks that everyone is capable of being turned from the shadows to the shadow-casting artifacts—of moving from the lowest segment (segment a) of the Divided Line to the next highest (segment b), the realm of belief proper.

Screens—television screens, phone screens, computer screens—are the Cave walls of today. When we uncritically accept the words and images we see there, we are like the chained prisoners. But if we turn and look at the sources of the information flickering before us, we might recognize that the information is distorted by bias and ulterior motive. Unlike the Cave's puppeteers, who do not seem to derive any benefit from their shadow-casting, the shadow-casters of our age typically *do* derive some benefit, and frequently their power depends upon our remaining chained, accepting the images they project before us, and believing that 'the truth is nothing other than the shadows' (7.515e). While being turned around is good for us, we often do not initially like it. But there is also a danger that in being turned around we will reject information we disagree with and take its source to be biased. Clearly, many sources *are* biased, but if we reject every artifact that comes from a puppeteer we do not like, it is not clear that we are any better off than we were before we turned to look. In fact, we might be worse off if we fall prey to the belief that critical thinking involves (merely) rejecting—perhaps as 'fake news'—anything emanating from sources we identify as 'liberal' or 'conservative' or whatever. Education, in the end, is not just any kind of turning around; it requires that the student be 'turned the right way' (7.518).

In the third stage we again see the role that compulsion plays in the Cave Allegory, for an unnamed, unidentified someone will 'drag [the freed prisoner] away from there by force, up the rough, steep path' (7.515e). Although Socrates devotes just one sentence to the third stage, what he says later in Book VII indicates that this rough, steep path symbolizes the formal education that potential philosopher-rulers receive. This four-subject education is the basis of the *quadrivium* of classical liberal education, the sort of education suitable to a free person. It is education centered on number: arithmetic (number itself), geometry (number in space), harmonics or music theory (number in time), and physics or astronomy (number in space and time). All these number-based subjects 'lead the soul and turn it around towards the study of that which is' (7.524e), which ultimately is the Form of the good. While there are certainly practical applications of these subjects, the would-be philosopher-queens and -kings study them 'not like tradesmen and retailers [... but] for ease in turning the soul around, away from becoming and towards truth and being' (7.525c). These disciplines prepare would-be philosophers not for craft-based careers in the sensible world, where they might be bakers or cobblers or doctors (although it will prepare them to be generals, as they are the city's guardians), but rather for citizenship in the intelligible world. They will learn to think abstractly, grasping essences and integrating Forms, which is presumably why studying geometry 'tends to make it easier to see the Form of the good' (7.526d). As the way out of the Cave, these subjects are 'merely preludes to [...] the song that dialectic sings' (7.531d–32a), and that is a tune that is sung only in the intellectual sunlight of the intelligible world outside of the Cave.

In stage four, the prisoner is not just freed from their fetters but has made it out of the Cave into the intelligible world above, which corresponds to the top half of the Divided Line (segments c and d). Looking at the fire in the cave hurt their eyes, and they find emerging into the sunlight painful, just as a mid-afternoon moviegoer who leaves a dark theater is pained by the bright parking lot outside. At first, they will only be able to look at shadows of the objects in the world above, here cast by the light of the sun rather than the fire, or their reflections in water, or look at the objects at night. Just as the shadows on the cave wall were mere copies of the artifacts held before the fire, those artifacts are

mere copies of the Forms, which are 'the things themselves' (7.516a). Although Socrates does not say, we can assume that there is one Form for each of the many particular objects in the cave. Whether there is one tree—the Form of treeness itself—or one oak tree, one maple tree, one white pine, one yellow pine, etc. is an interesting question to ponder, but it is not one we need to answer to understand the Cave Allegory or the *Republic* as a whole.

At stage five, the former cave dweller is able to look directly at the sun, 'not images of it in water or some alien place, but the sun itself, in its own place, and be able to study it' (7.516b). Presumably not everyone who makes it out of the Cave is able to do this. Mathematicians and scientists study the Forms relevant to their disciplines, but they do not see other Forms or how the Forms they contemplate are related to these other Forms, and they certainly do not see the Form of the good—that vision is reserved for genuine philosophers, and there are *very* few of them. So presumably the fourth stage in the Cave Allegory corresponds to Thought on the Divided Line, while the fifth stage is where Understanding operates.

In stage six, the sun-contemplating philosopher first thinks back on his life in the cave, and reflecting on 'what passed for wisdom there' (7.516c), smiles ruefully and feels pity for the others still trapped in their ignorance, who 'know' only the shadows on the wall or the artifacts casting them. What would happen if the enlightened philosopher descended into the cave? They will not be greeted as a returning, liberating hero, Socrates thinks. The denizens of the dark world below will first think the returning philosopher a fool: until their eyes, used to the bright light of the intelligible world, have adjusted to the darkness of the cave, they will be unable to recognize the shadows or the puppets. Like the ship owner who thinks the true captain is a useless stargazer (6.489c), the cave dwellers will think the enlightened philosopher a fool who has ruined their eyesight (not to mention his economic prospects) by looking too long at the sun. But if they persist and try to free the prisoners and turn them toward the firelight or drag those who are able out of the cave, they will think their 'liberator' is worse than useless: they will think them dangerous, and 'if they could somehow get their hands on him [...] they [would] kill him' (7.517a).

Plato does his readers a good turn by having Socrates explicitly connect the Sun and Cave metaphors (7.5157bc), but he leaves the task of fitting together the Divided Line and Cave to us. Fortunately, connecting them is fairly straightforward, as we have already seen. The shadows on the Cave's wall correspond to the images seen at the Divided Line's lowest section (segment a), the realm of *Imaging*. The shadow-casting puppets held before the fire correspond to 'the originals of [the Line's] images' (6.510a), in segment b. Just as the shadows are copies of the originating artifacts, these artifacts, which are at home in the Visible World, are in turn copies of the Forms, which of course reside in the Intelligible World (section c). Plato is not suggesting that the images, shadows, and reflections are not real, but rather that they are *less* real than the originals they are images of. This has a lot of intuitive appeal: I can create a shadow of my hand by interposing it between my desk and lamp, but the shadow cast seems less real than my hand in at least a couple of ways. First, while my hand is a three-dimensional object, the shadow is only two-dimensional, lacking the dimension of depth. Second, the shadow depends for its existence on the presence of my hand (and on the presence of the 'third thing' that features in the Sun Analogy: light). My hand still exists when I turn off my desk lamp or move it out of the lamp's range, but the shadow no longer exists. Shadows, reflections in mirrors and water, etc.—the stuff of segment a of the Line—are ephemeral. They are not unreal—my seeing the shadow is not an optical illusion: there is something there, just something whose existence is thinner and flimsier than the objects at the Line's second section (segment b). Now—and here's the metaphysically important point—just as the shadows and reflections are copies of what seem to be independently existing objects, these objects themselves are copies of the Forms they instantiate. The bed the carpenter makes, Socrates argues in Book X, is 'something which is like that which is' (10.597a). The second 'is' is the 'is' of existence: the built bed is like what is real, what fully exists. Its resembling the Form of bedness is what makes it a bed and not a table, but, just as Van Gogh's paintings of his bed at Arles are copies of the bed he slept in, so too is that bed a copy of the Form. Thus Plato's metaphysical point can be put as a ratio, image: original :: original: Form.

There is much that Plato leaves unsaid about the Cave. Who first frees the prisoner? Who drags them up and out of the cave? Who are the puppeteers? What, if any, benefit do they derive from keeping the prisoners occupied with shadows? Glaucon says toward the outset that these are 'strange prisoners', to which Socrates replies, 'they are like us' (7.515a), so with a bit of imagination we can fill in some of these blank spots.

Trouble in Paradise: The Powers Argument Casts a Shadow on the Cave Allegory

There is a problem lurking in the background of the Cave Allegory that should be brought to the forefront and addressed. Indeed, all three of the key analogies—the Sun, the Divided Line, and the Cave—are analogical or metaphorical accounts of two distinct worlds or realms: the intelligible world, where the Forms reside, and the visible world, home to spatiotemporal particulars. The Powers Argument was supposed to provide some reason for believing in Plato's two-worlds metaphysics and indeed for taking the Metaphysical Elevator to the fourth floor, where the Forms are not just real but are *more* real than the particulars that instantiate them. But we have seen that the Powers Argument is logically invalid, since its conclusions could be false even if its premises are true, and even if its logical problems could be fixed, it would still not provide a good reason to accept Plato's two-worlds metaphysics, given the implausibility of belief and knowledge having distinct objects. Presumably what Socrates says about agreed-to hypotheses that prove to be false goes for arguments, as well: 'that if it should ever be shown to be incorrect, all the consequences we have drawn from it will also be lost' (4.437a). Since the Sun, Divided Line, and Cave Analogies all require the distinction between the Visible and the Intelligible Worlds, they are infected, perhaps fatally, by the failure of the Powers Argument.

Of course, Plato's two-world metaphysics could still be correct, since the conclusion of an unsound argument can still be true. But unsound arguments do not justify belief in their conclusions. What should we make of the major analogies of Books VI and VII in light of the failure of the Powers Argument? One option is to proceed in a hypothetical or conditional way: *if* these are the two worlds, then here is how they differ.

Given the hypothetical nature of Socrates' procedure in the *Republic*, this is not a bad way to go. Another option is to interpret the two worlds non-literally but metaphorically, which fits well with the prevalence of metaphor in the *Republic*. On this view, the two worlds are ways of thinking about or conceptualizing reality rather than assertions about the nature of reality itself. And perhaps this is how the two-worlds metaphysics should be interpreted even if the Powers Argument were sound: think of the Forms populating the intelligible realm (and that realm itself) as useful fictions. This metaphysically more cautious view would appeal to fans of Ockham's Razor.

A problem with the metaphorical interpretation, however, is that Plato himself seems to take the two worlds literally: 'there are these two things [i.e., the Form of the good and the sun], one sovereign of the intelligible kind and place, the other of the visible' (6.509d). Plato thinks of these as *places*, which suggests their reality. Although the word being translated as 'place' (τόπος [*topos*], whence the English word 'topographical') could mean *realm* in a non-physical sense, it is difficult to think that Plato intends his talk of the Forms and the intelligible realm to be taken only metaphorically. Still, for readers bothered by the failure of the Powers Argument, this may be the best interpretation, even if it is not Plato's. After all, we can still distinguish Understanding from Thought, the two kinds of cognition at work in the intelligible realm, without being realists about the Forms. Even if we take Plato's Metaphysical Elevator only to the second floor, we can still distinguish people who grasp the essence of a perhaps narrow range of things from people who do not merely grasp more essences but also *see connections* between them. Integrative thinking is one of the hallmarks of dialectic, and one can prize that capacity while at the same time denying that the Forms exist mind-independently. Even Socrates himself is agnostic—in the literal sense of not knowing—about the metaphysical status of the Forms and the intelligible realm: 'Whether it's true or not, only the god knows' (7.517b). He seems to *believe* that the Forms are real, but perhaps this remark is Plato's way of indicating that he is aware of the Powers Argument's shortcomings: Socrates himself does not think he has proven the argument's conclusion. Early in Book X he recounts his 'usual procedure', which is to 'hypothesize a single form in connection with each of the many things to which we apply the same name' (10.596a).

So he seems aware of the hypothetical, mathematician-like nature of his investigation. If Socrates can live with this sort of uncertainty, perhaps readers can as well.

Food for thought. But it is now time to turn to another worry about the Cave Allegory, the enlightened philosopher's return to the world below.

Going Back Down into the Cave (7.519b–520b)

Having been fully liberated from the dark, smoky world of the Cave, the enlightened philosophers are in no hurry to return. Readers will remember that the problem Socrates faces in responding to the Third Wave—and it is worth remembering that the Sun, Divided Line, and Cave analogies are all part of that response—is that in the actual world, ‘political power and philosophy’ are separated, with philosophers as uninterested in participating in the messy world of politics and government as those in power are in studying metaphysics and epistemology. The solution, both to the ideal city's real possibility and to individual and communal happiness, is that these philosophers and political leaders be ‘forcibly prevented’ from pursuing their own interests exclusively. Somehow, ‘political power and philosophy [must be made to] entirely coincide’ (5.473cd). The philosophers would prefer to remain in the sunlit world above, contemplating the Forms. But possessing knowledge of the good, they and they alone are capable of governing. They have been compelled to ascend to the sunlit, intelligible world above; is it fair to compel them to go back down to the dark, smoky cave, the visible world below?

There is a substantive philosophical problem for Socrates' view that the just life is happier than the unjust life, an issue that Plato does not notice or at least does not remark upon. But before investigating that, we should attend briefly to one of the *Republic's* most gratifying literary delights. The issue before us is compelling the enlightened philosophers to go back down into the cave to govern it. The Greek word at issue is καταβαίνειν (*katabainein*), to go down. If you turn to the *Republic's* first page, you will see that Socrates' first words, the very first words of the *Republic*, are ‘I went down’ (1.327a). The Greek there is κατέβην (*katebên*), the first-person singular form of καταβαίνειν in the aorist

(past) tense. The implication is that we are all cave-dwellers and that Socrates' going down to the Piraeus is like the enlightened philosopher's going back down into the cave, where we muck about in the dark as we look for justice. It is no wonder that, having discovered the other three political virtues (wisdom, courage, moderation), Socrates finds justice hard to locate at first: 'the place seems to be impenetrable and full of shadows [...] dark and hard to search' (4.432c). The conversation that is the *Republic*, then, takes place in the Cave, where 'we contend about the shadows of justice or the statues of which they are the shadows' (7.517d). Although Plato could have had Socrates just say this simply and directly, it is more powerful and more aesthetically pleasing for readers to see this for themselves. Although some readers will yawn, others will be delighted at Plato's literary artistry, and perhaps will be able to understand more fully why some people devote their lives to understanding and appreciating his philosophical thought and literary craft and the way he integrates them.

Now on to the substantive philosophical question of the enlightened philosopher's return to the cave. An important point to grasp is that the liberated philosopher is not on a mission of liberation, at least not complete liberation, since on Plato's view not everyone is capable of making it out of the cave. As we have noted several times already, he thinks that 'the majority cannot be philosophic' (6.494a). The returning enlightened philosopher will free whom he can, dragging those who are able to follow 'up the rough, steep path' (7.515e), but their main task is to govern in the Cave—'to guard and care for the others' (7.520a). We know they will not be received well, but if through 'some chance event' or divine intervention (6.499b) they are able to take charge of the cave, they will govern well, since they have the virtue needed to do so: political wisdom. Even so, the philosophers do *not* want to return to the Cave, and interestingly enough, Plato takes this as a plus: 'A city whose prospective rulers are least eager to rule must of necessity be most free from civil war, whereas a city with the opposite kind of rulers is governed in the opposite way' (7.520d). Since 'it is those who are not lovers of ruling who must rule' (7.521b), the returning philosophers' reluctance counts in favor of their doing so.

So why do the philosophers descend into the cave and do what they do not really want to do? Just as they were compelled to ascend out of the

cave, they are compelled to descend into it, but the compulsion in the two cases is different. They were physically dragged up, at least metaphorically speaking, but they are not physically dragged back down. It is no accident that their being compelled upward would be metaphorically physical, since our particular individual bodies belong in the visible realm of particulars, while our souls, by contrast, are not physical, so there is nothing to drag. Their being compelled downward is mental or psychic, but it is not the irrational or non-rational compulsion that consists in brainwashing or advertising by people who seek to *cause* us to pursue ends they have chosen for us. Instead, the enlightened philosopher is compelled to return by rational persuasion. If one recognizes that an argument is sound—that its conclusion must be true if its premises are true *and* that its premises are in fact true—one is rationally compelled to accept the conclusion. This is not as a matter of internal or external causation, but rather of rational compulsion: the force of rational persuasion. So what is the argument that the enlightened philosophers should find so compelling?

As I hinted earlier, Socrates does not argue that any enlightened philosopher has a duty to descend to the cave and govern. Instead, only those whose enlightenment results from the city's having educated them have that duty. 'What grows of its own accord and owes no debt for its upbringing', he argues, 'has justice on its side when it is not keen to pay anyone for that upbringing' (7.520b), where the currency of repayment is governing. It might be noble of an accidentally- or divinely-self-enlightened philosopher to do what they do not really want to do and descend into the cave, but this is not a requirement of justice: they would not act wrongly were they to remain above, pursuing their philosophical interests. But a philosopher educated by the city has a duty of reciprocity and gratitude to descend and govern.

We should note that Plato here shows that he is not a consequentialist about morality. We can assume that the consequences of the philosopher's descending would be better, all things considered, than the consequences of their remaining in the intelligible world above. If we take the good they would do by governing, which is presumably substantial, since there can be no real happiness for the citizens if philosophers do not rule (5.473e), and subtract from it the personal cost to them of sacrificing their own preferences for the good of the group, the net consequences of descending would still be overall better than those of not descending.

But this fact alone is not sufficient to generate a duty for the enlightened philosopher to descend and govern. Were they to go down into the cave, they would be going beyond the call of duty—going down would be supererogatory, as philosophers say. Supererogatory actions are praiseworthy to perform, but not blameworthy to omit. Donating a kidney to a stranger is, other things being equal, praiseworthy, but my not doing this is not blameworthy: I do not act unjustly if I keep both of my kidneys. (Of course, if I have promised to donate the kidney and the stranger has relied on my promise, then 'other things' are not equal, and the moral situation has changed considerably.) Actions required by justice are different: failure to perform them *is* blameworthy, and, other things being equal, performing them is not praiseworthy. Special circumstances are required for refraining from violence to be praiseworthy, as envisioned in a *Sopranos* episode when Tony, a violence-prone mafioso, forgoes killing his daughter's sexually predatory soccer coach and lets the police deal with it. 'I didn't hurt nobody today', Tony drunkenly tells his wife, and for him, this is suitably praiseworthy.¹ For the rest of us, not killing people who bother us or whom we regard as moral reprobates is what is minimally expected, and there is no praise for doing what we ought to be doing.

So not just any enlightened philosopher, but only the enlightened philosopher who owes their enlightenment to the education that the city has provided for them, has a duty to go down into the cave and govern. Here is the argument Socrates gives them:

We have made you kings in our city and leaders of the swarm, as it were, both for yourselves and for the rest of the city. You are better and more completely educated than the others and are better able to share in both types of life. Therefore each of you must go down to live in the common dwelling place of the others and grow accustomed to seeing in the dark. When you are used to it, you'll see vastly better than the people there. And because you have seen the truth about fine, just, and good things, you'll know each image for what it is and also that of which it is the image. Thus, for you and for us, the city will be governed, not like the majority of cities nowadays, by people who fight over shadows and struggle against one another in order to rule [...] but by people who are awake rather than dreaming (7.520bc)

1 *The Sopranos*, Season 1, Episode 9, 'Boca', dir. by Andy Wolk (HBO, 1999).

This is an interesting argument, and it certainly has a lot of intuitive appeal. There is something compelling, after all, about obligations of gratitude: if you have gone out of your way to benefit me, I seem to incur a debt of gratitude. What it takes to repay that debt varies with the circumstances: often, a simple 'thank you!' is all that is required, but other times—as in the present case—more is required. Here, Socrates argues that, as a matter of justice, the enlightened philosophers must (temporarily, at least) give up the life they prefer—a philosophical life devoted to contemplating the Forms—for a life of political action. (These are the 'both types of life' referred to in the quotation above.)

Plausible though the argument is, there is something troubling about duties of gratitude, even when the benefit to be reciprocated was bestowed intentionally, for the sake of the beneficiary. The worry is that one can go around obligating others to do good turns for oneself by doing good turns for them. If I show up unbidden and start harvesting your wheat for you, does my supererogatory act really bind you to do the same for me? Many of us would *feel* obligated to reciprocate, but the issue is not the psychological one about our feelings but rather the philosophical, normative one about our duties. Consider how your views would change if the helpful harvester helped not primarily because he wanted to benefit you, but because he needed your help harvesting his large wheat field, and, knowing you to be a 'nice' person but not wanting to ask for your help, decided that the best way to get you to help him was to help you. I suspect you would feel a bit manipulated. And suppose that your neighbor harvested your wheat when you were away in town on Saturday, without asking if you needed or wanted their help. They would have imposed this benefit on you, without your consent. Your supposed duty to return the favor would look flimsier and flimsier.

The trouble with Socrates' argument is that the city's actions in educating the philosopher too closely resemble the 'helpful' neighbor harvesting your wheat. Remember that the would-be rulers are *compelled* to leave the cave: 'someone dragged him away from there by force, up the rough, steep path, and did not let him go until he had dragged him into the sunlight' (7.515e). The benefit has been bestowed and received non-voluntarily, which surely makes a difference to whether there is a duty of gratitude to reciprocate. In addition, Socrates misspeaks when he claims that the philosophers were educated 'both for [them]selves

and for the rest of the city' (7.520b). Any benefit the philosophers personally receive is foreseen, but not intended. Given the strong communitarian thrust of the ideal city, it is clear that the education is not primarily intended for the philosopher's benefit but rather for the city's; any benefit the individual philosopher receives is a side effect or by-product.

But even if Socrates' argument for a duty to return is sound, there are disquieting implications for his view that the just life is happier than the unjust life, that justice benefits its possessor. Thrasymachus insisted that while justice benefits others it is always bad for its possessor: being just and acting justly makes one worse off in the long run. Thus Thrasymachus sees a wedge between *what is good* (or right) and *what is good for me*. Consider the situation of the enlightened philosopher. They would strongly prefer to remain in the Intelligible World, basking in its sunlight and contemplating the Forms. They will return because, being just, they will do what justice requires of them, even when they do not want to do it. But make no mistake about it, they do not want to return, and ruling is 'something compulsory' (7.520e), not enjoyable in itself as doing philosophy is. And notice that ruling does not really fit into any of the three categories of goodness that Glaucon articulates at the beginning of Book II. While ruling seems at first to belong to the category of goods that are 'onerous but beneficial' (2.357c), upon reflection we can see that it does not really fit there, since this mixed category contains goods that are 'onerous but beneficial *to us*' (my emphasis). Few people enjoy flossing their teeth, but those who do this regularly derive a benefit and presumably decide that on the whole flossing is worth it: but its value is extrinsic and instrumental, not intrinsic. But imagine if flossing benefited not the flosser but someone else. This seems to be the position of the enlightened philosopher. They return to the cave to govern, but they would rather not, since they would be personally better off ignoring the demands of justice. When Glaucon worries that justice is 'making them live a worse life when they could live a better one' (7.519d), Socrates does *not* reply that they *are* better off acting justly; instead, he reprises the response he made to Adeimantus at the beginning of Book IV: his concern is not 'to make any class [or particular citizens]

in the city outstandingly happy but to contrive to spread happiness throughout the city' (7.519e; compare 4.419a and 5.466a), which concedes the assessment underlying Glaucon's question.

The philosopher's return benefits the cave's residents, since 'there can be no happiness, either public or private' in any city not governed by a philosopher-king or -queen (5.473e). But returning does not benefit them personally, and that is the real issue here. Socrates seems to be conceding that Thrasymachus is right after all: justice benefits someone else, not its possessor. Even if the overall consequences of the philosopher's returning were better than the consequences of their remaining above, their return would not benefit them. So it looks like justice does *not* benefit its possessor: leading a good life seems to come at the cost of having a good life.

Thus it seems that the philosopher's situation is analogous to the far-fetched scenario in which flossing does not benefit the flosser but somehow benefits others. It would be a mistake to think that even in this scenario I would have no self-interested reasons for flossing. If there is a community norm that everyone should floss, my flossing would help sustain and promote this norm (and thus indirectly contribute to the benefit adherence to the norm produces) and encourage others to do so as well. If this is the case, then my flossing would benefit me indirectly. In doing my part to uphold norms that benefit the community, the burden of compliance might be counter-balanced by the benefit received. The same might be said for the returning philosopher, who lives a better life in a well-governed city than they do in the poorly governed city of the Shelter from the Storm analogy, which we considered in the last chapter. Although they would rather not descend, perhaps the philosopher's doing so really does benefit them when we look at the big picture. So perhaps Socrates does not give away the game to Thrasymachus after all. Even though the philosopher's return seems altruistic—they return to 'labor in politics and rule for the city's sake' (7.540b) rather than their own, they might in fact benefit by their return. Though the city's good is the outcome they intend, they can perhaps foresee that they will benefit too.

A worry remains, though: in the imaginary scenario in which flossing benefits others, it seems unlikely that *my* not flossing will have bad

consequences so long as enough of my fellow citizens floss regularly. Or, to take a less far-fetched example, I might reason that while I enjoy National Public Radio, I can still receive this benefit without bearing my share of the burden, since NPR's not receiving \$100 from me will not cause them to close up shop. It seems to be in my self-interest to be a free-rider, benefiting from the good behavior of others while not burdening myself with doing my share. The ethics of Immanuel Kant rules out such free-riding behavior: if everyone's acting on the maxim or principle I plan to act on would make it impossible for me to act on it, then my acting on it is wrong. But Kant did not share Socrates' view that doing the right thing makes me better off all things considered: the demands of morality are frequently at odds with those of self-interest and happiness.

Now perhaps free-riding would not even tempt the fully just philosopher, who takes their turn at ruling without complaint. But the self-interested Thrasymachan, who is 'vicious but clever' (7.519a), is unlikely to be persuaded: the philosopher would clearly be better off if they missed a turn every once in a while, if they called in sick when they really wanted a day of metaphysical sun-bathing. And things look even worse for the view that the just life is happier if we bear in mind the lives Socrates is to compare to settle the question of which life is happier: a just person who appears unjust versus an unjust person who appears just. The philosopher who does not go back down to the cave would be unjust, but under the terms agreed to they would not appear to be so: their free-riding would have to go unnoticed and thus would not undermine the norms governing the small community of philosophers, so their not going back down to the cave to rule benefits them without the negative effect on norms of justice.

These are some of the issues readers will want to keep in mind as we explore Books VIII and IX, where Socrates resumes his investigation of the *Republic's* second question. In Book VII, though, he does not seem to notice them—or if he does, he gives no explicit indication of this.

Some Suggestions for Further Reading

There is a large literature on Plato's Allegory of the Cave. Readers interested in the thought of Martin Heidegger will want to see *The Essence of Truth*, trans. by Ted Sadler (New York: Continuum Books, 2002). For an account more in keeping with the style and concerns of contemporary Anglophone philosophy, readers might turn to Chapter 10 (Understanding the Good: Sun, Line, and Cave') of Julia Annas, *An Introduction to Plato's Republic* (New York: Oxford University Press, 1981), pp. 243–71, reprinted in *Plato's Republic: Critical Essays*, ed. by Richard Kraut (Lanham, MD: Rowman & Littlefield, 1997), pp. 143–68.

Readers with a taste for serious cinema and an interest in the Cave Allegory will certainly want to watch *The Conformist*, dir. by Bernardo Bertolucci (Paramount Pictures, 1970), about a young fascist tasked with assassinating his former philosophy professor. The film is rife with Platonic imagery as well as a cinematically brilliant discussion of the Cave.

Interested readers can find an animated version of the Cave Allegory on YouTube, narrated by the great Orson Welles, at https://www.youtube.com/watch?v=_jmJGBJRIUQ.

Readers interested in the enlightened philosopher's descent back into the Cave should see Richard Kraut, 'Return to the Cave: *Republic* 519–521', in *Plato 2: Ethics, Politics, Religion, and the Soul*, ed. by Gail Fine (New York: Oxford University Press, 1999), pp. 235–54.

Readers interested in gratitude as a basis for duties of justice might start with Chapter 7 of A. John Simmons, *Moral Principles and Political Obligations* (Princeton: Princeton University Press, 1979), pp. 156–90.