The Unity of the Soul in Plato's *Republic*

Eric Brown  
Department of Philosophy  
Washington University in St. Louis  
eabrown@wustl.edu

1. Introduction

In Book Four of Plato's *Republic*, Socrates divides the soul.¹ He argues that because one thing cannot do or undergo opposites in the same respect, in the same relation, and at the same time, some cases of psychological conflict can only be explained by supposing that the soul is actually more than one thing. But Socrates does not deny that the soul is a unity. Rather, he holds that the soul, like so many entities, is both a many and a one. He refers to the divisions as parts of the soul,² and even when he ascribes actions and attitudes to a part of the soul, he also ascribes them to the whole soul (esp. 439a-b, 439d). Although few souls enjoy the unity that virtue earns,³ every human soul possesses an unearned unity.⁴

---

¹ I thank the organizers of the Toronto conference and the editors for including me. I also thank Jennifer Whiting for the paper on which I commented in Toronto and Devin Henry for the opportunity to present a successor to those comments at the University of Western Ontario. Finally, I thank those who offered helpful comments on the successor, especially Verity Harte and John Mouracade in Ontario, and Clerk Shaw, Rachel Singpurwalla, Eric Wiland, and the editors by correspondence.

² For the *Republic*, I cite the text of Slings 2003, and translations are mine, though I have borrowed heavily from Grube's translation, as revised by Reeve, which appears in Cooper 1997.

³ Socrates calls the divisions εἴδη, γένη, and μέγη: εἴδη in 435c1, 435c5, 435e1, 439e1, γένη in 441c6, 443d3, μέγη in 442b10, 442c4, 444b3 (cf. 577d4, 581a6, 583a1, 586e5). He also refers to the whole soul in 436b2 (cf. οὐνόματον at 611b5-7).

⁴ For earned unity, see especially 443d-e. That it is rare follows from these three points: virtue requires knowledge (442c), knowledge requires philosophy (474b-480a), and philosophy is rare (491a-b).

⁵ I discuss an apparent exception (554d) in the last section.
What explains this unearned unity? How do reason, spirit, and appetite constitute a single whole? This question has received little attention.\(^5\) My first aim here is to establish that it is an important puzzle, and my second is to venture an answer.

2. The importance of the soul’s unity

My question is in part mereological. Plato assumes that my reason, spirit, and appetite constitute a complex unity, and that my reason, spirit, and appetite together with your spirit do not. What explains which psychological elements constitute a whole soul and which do not?

The interest of my question, however, is not exclusively mereological, and Socrates’ talk of “parts” and “wholes” does not exhaust the reasons for asking it. Plato needs to explain the unearned unity of the soul. First, if the soul is, for all that he can explain, three distinct entities, then he cannot account for the felt unity of consciousness.\(^6\) As Descartes puts it, “When I consider the mind, or myself insofar as I am merely a thinking thing, I am unable to distinguish any parts within myself; I understand myself to be something quite single and complete.”\(^7\) In the *Republic*, Socrates does not address this unity of consciousness directly, but he assumes that a human agent is somehow aware of what every part of his or her soul thinks and feels. Unless the human agent is something over and beyond the parts, it is somehow composed of the parts. How?

Second, if Plato cannot explain the unity of the soul, then he cannot treat the whole soul as the locus of moral responsibility. Socrates speaks as though our ordinary practices of praise and blame correctly take the same agent to be responsible whether he pursues wisdom, reacts angrily, or seeks a yummy dinner (cf. 436a8-b4, quoted

---

\(^5\) But see Aristotle *De an.*, I 5 411b5-7, Archer-Hind 1882, 124, and Lorenz 2006, 38-40. Some scholars offer a principled excuse: they maintain that the soul is a simple unity with three different kinds of psychological states and activities that are mere conceptual parts (e.g., Shields 2001). I argue against this reading in §3 below.


\(^7\) *Meditation Six*, in Cottingham, Stoothoff, and Murdoch 1984, 2:59, rendering the text in Adam and Tannery 1964-1976, 7:86.
below), even if that person's soul is tyrannically divided against itself (572b-580a, esp. 577d13-e1). Plato cannot mean that the complex of soul and body is the proper locus of responsibility, since the Myth of Er makes it clear that the disembodied soul in the afterlife is responsible for what it did while embodied (614b-616b).\(^8\) (Besides, the soul is supposed to be responsible for what the body does [353d].) It might be tempting to make the rational part alone responsible, since Socrates suggests that only the rational part of the soul survives separation from the body (611b-612a)\(^9\) and Er’s myth says that the soul in the afterlife bears responsibility for the person’s past embodied existence (614b-616b) and for the next (617d-e). But many actions for which an embodied human being is responsible are motivated by appetitive or spirited desire, and holding the rational part responsible for such actions seems to blame an innocent bystander or, worse, a victim. So it is better to infer from Socrates' manner of speaking that the whole soul is responsible. This can accommodate the hypothesis that only the rational part of the soul survives death. Plato could take the whole complex soul to be the locus of moral responsibility while it is embodied and the whole remaining (incomposite) soul to be the locus in the afterlife. He would be justified in doing this because these are the same soul. (In connection with the body, the soul develops parts that it lacks without a body [cf. 518d9-519a1], but it is nonetheless the same entity with or without the body.) He could elaborate this justification if (as I will suggest) what explains the complex unity of the embodied soul also explains how reason (and each of the other parts) is at least indirectly responsible for everything that any part of the embodied soul does.

There are, then, good reasons to take Socrates' talk of parts and wholes seriously and to seek an explanation of how the parts of every soul compose a single whole. But these reasons are not necessarily conclusive. First, if Plato's best account of unity does not attribute unity to every soul, one might take another approach to all

---

\(^8\) For the earned unity of the soul and body, see 462c-e. Socrates does not discuss the unearned unity of this complex, though this is surely relevant to the unity of consciousness (think of sense-perception) and to moral responsibility.

\(^9\) Socrates is not fully committed to this suggestion (612a3-6), but he plainly considers it plausible. See also 518d9-519a1, the *Phaedo* (esp. 78b-84b), and the *Timaeus*, but contrast the *Phaedrus*, *contra* Archer-Hind 1882, 127.
this evidence. One might, for example, say that Socrates speaks falsely of parts and wholes and of single agents, for merely practical reasons. Second, even if Plato can account for unity in every soul, there might be evidence that calls into question Plato's commitment to the unearned unity of every soul. So I aim to show that Plato can account for the unity of every soul, and that there is no evidence undermining his account.

There remain two other challenges to the claim that the soul's unearned unity is a significant puzzle. One might think that Plato faces no real puzzle because he conceives of every soul as a simple unity, with merely conceptual parts, like subsets of psychological attitudes or "aspects" of a single entity. Alternatively, one might think that Socrates' way of dividing the soul simply rules out its unearned unity, so that his talk of a "whole" soul comprising "parts" and his endorsement of ordinary practices of praise and blame and of the unity of consciousness call for an error theory or a fresh start in another dialogue. These views assume that if the Republic assigns three distinct subjects to the soul, it cannot account for the soul's unity. The first performs modus tollens with this conditional, the second modus ponens. I reject the conditional itself. First, I show that according to Book Four's division of the soul, the soul's unearned unity must and can be complex. Then I will identify a Platonic account of complex unity that explains why the whole soul is a locus of moral responsibility and at least enables explaining the unity of consciousness.

---

10 Cf. note 36 below, concerning Harte 2002.

11 Cf. Kamtekar 2006, minus her insistence that Plato would not find protreptic value in a wholly false account.

12 See, e.g., Shields 2001 and (though he does not go for this approach wholeheartedly) Price 2005 and 2009.

13 See Grote 1888, 2:386n ("three souls"); Adam 1905, ad 435a, qualified by the note ad 439b; and Stalley 1975. Bobonich (2002, 219-257) has recently argued forcefully for this conclusion, though he also allows himself to talk about the composite soul without explaining the composition or admitting that it offers any unity of the person. His cogent denial (234 with 531n27) that there is any ultimate subject above and beyond the three parts of the soul does not exclude the possibility that the three parts might compose a unified subject.
The problem of complex unity

The puzzle of how Plato might account for the unearned unity of the soul stems from Socrates' main argument to divide it. Socrates introduces the argument by suggesting that there might be three different things in us "by which" we do three different kinds of things. He says,

But this is hard, [to determine] whether we do each of these things by this same thing or by three things, one thing by one and another by another. Do we learn by one thing, get angry by another of the things in us, and again by a third desire the pleasures concerning nutrition and procreation and all the things akin to these, or do we act in each of these cases by the whole soul, whenever we should have an impulse to act? This is hard to determine in a way worthy of our argument. (436a8-b4)

Socrates and Glaucon then agree to "try to determine the things in us, whether they are same as each other or different" (436b6-7).

This opening raises the problem of the soul's unity without taking a stand. It plainly introduces the possibility that there are three distinct sources of psychological states and activities in us. Socrates is not merely categorizing the kinds of psychological states and activities; he has already done that easy task (435d-436a). He now raises the hard question of whether we should attribute these different kinds to different causal sources in us. This already tells against the suggestion that the soul has merely conceptual parts. But it does not rule out the possibility that the distinct

---

14 Socrates says "in a way worthy of our argument" not to say that their standards are especially high but to remind his interlocutors that they are not as high as they might be: they are not traveling the "longer road" that leads to a fuller answer by means of the Good (435c9-d3 with 504b1-7). Socrates continues to remind his interlocutors of these lower standards when he notes that the principle of non-opposition is merely a hypothesis in their current argument (437a3-8). I take these to be related because I assume that the Good is supposed to be the unhypothetical first principle (see 533a-534c with 505a, 506d-509d, and 509d-511d). Hence, I take him to be saying that the longer argument would be required to render their conclusion in Book Four unhypothetical. If that is right, Socrates does not suggest that the longer road would necessarily lead to a substantially different conclusion.
causal sources could compose a complex unity. Socrates and Glaucon simply do not consider the ramifications of there being different sources for learning, anger, and desires for bodily pleasures. For all they say, the different sources might be parts of a single soul, as they are all sources "in us," or they might fail to be parts of a single soul, despite their being "in us."

To answer his question, Socrates introduces the principle of non-opposition, a general principle applicable to any entity whatsoever: "It is clear that the same thing will not be willing to do or undergo opposites [1] in the same respect [κατὰ ταῦταν γε] and [2] in relation to the same thing [πρὸς ταῦταν] [3] at the same time [ἄμα]" (436b9-10). He immediately draws out the consequence for the "things in us" that he has mentioned: "The result is that if we should somehow discover these things [viz., opposites] happening in them [viz., the things in us by which we learn, feel anger, and desire the pleasures of nutrition and generation and such], we will know that they were not the same thing, but many." Socrates continues to presuppose that either we act by one thing or we act by many things. So either he continues to set aside the ramifications of acting from multiple psychological sources—perhaps the multiple psychological sources are parts of a single soul and perhaps they are not—or he assumes without argument that the only way the soul can be unified is for its apparently

---

15 It might seem to. Socrates contrasts acting by one of three psychological sources with acting "by the whole soul [ὅλῃ τῇ ψυχῇ]" (436b2). But if this were the only way to understand the possibility of acting by the whole soul, then how would Socrates explain the implication of the whole passage (436a8-b4) that a single human agent acts whether this or that part of her soul is the source of the action? We need to distinguish acting by the whole soul underderivatively (which contrasts with acting with one of three psychological sources) and acting by the whole soul derivatively, insofar as one is acting by one of three psychological sources. Socrates' question presupposes the first and ignores the second. The implication of the whole passage, to which Socrates returns (439c3-8), requires the second.


17 436b10-c2: ...ὡστε ἄν ποι ἐνορίσκωμεν ἐν αὐτοῖς ταῦτα γινόμενα, εἰσόμεθα ὅτι οὐ ταῦτά ἦν ἀλλὰ πλείω. My supplement takes αὐτοῖς to refer to the same things that the αὐτὰ of 436b6-7 refer to, namely the things in us discussed in 436a8-b4. Unfortunately, the Grube-Reeve translation (in Cooper 1997) is misleadingly careless here. It reads, "So, if we ever find this happening in the soul, we'll know that we aren't dealing with one thing but many." The careless "in the soul" for ἐν αὐτοῖς apparently leads to the complete fabrication of "we aren't dealing with one thing."
multiple sources of activity to be the same. He has explicitly said nothing to shut the
door on complex unity.

But what has he implied? The principle of non-opposition allows that a single
subject can undergo opposites (call them \( \varphi \) and \( \psi \)) only when at least one of three
conditions is met. A single subject can \( \varphi \) at one time and \( \psi \) at another; a single subject
can \( \varphi \) in relation to \( A \) and \( \psi \) in relation to \( B \); and a single subject can \( \varphi \) in one respect
and \( \psi \) in another. But can a single subject \( \varphi \) by one part and \( \psi \) by another? Or has
Socrates already ruled this out?

According to what I will call the standard reading, a single subject can \( \varphi \) by one
part and \( \psi \) by another, because to \( \varphi \) by one part and \( \psi \) by another just is to \( \varphi \) in one
respect and \( \psi \) in another. On this reading, the principle of non-opposition allows the
soul to be a whole that comprises parts that are distinct sources of psychological states
and activities.

Unfortunately, it is not hard to make trouble for the standard reading. Socrates
immediately clarifies the principle of non-opposition by considering two potential
counter-examples to it. First, he says,

If someone were to say of a human being who is standing still but moving his
hands and head that one and the same person is standing still and moving at the
same time, we would not, I think, think that we should say this, but rather that
one thing of him is standing still and another [thing of him] is moving. (436c10-d1)

Socrates does not say that the single human being stands still in one respect (by some
part of him) and move in another (by another part of him). He insists, instead, that one
thing stands still and another thing moves. Nothing prevents his point from

---

18 See, for examples, Irwin 1995, 204; Price 1995, 40-41; and Shields 2001, 145. The Grube-Reeve
translation (in Cooper 1997) unfortunately forces this reading on readers. They render the
principle of non-opposition's \( \kappa\alpha\tau\alpha \tau\omega\tau\omicron\omicron \) (436b9: "in the same respect") as "in the same part of
itself," and they use the same rendering for \( \kappa\alpha\tau\alpha \tau\alpha\omicron\omicron \) at 436c6.

19 I draw on Bobonich 2002 and Lorenz 2006 for the "trouble-making" reading. Both believe that
Socrates establishes three distinct "ultimate" or "proper" subjects of psychological states and
activities, though they disagree about the nature of the subjects, their oppositions, and the
implications of the division.
generalizing. If two parts of a single thing are subjects of opposites, the single thing
does not undergo opposites in different respects; rather, a many undergoes opposites.

This leaves unclear what it would mean to say that something $\varphi$s in one respect
and $\psi$s in another. If talk of different "respects" does not cover different parts, then what
does it cover? The trouble-makers take Socrates' treatment of the second potential
counter-example to answer this question:

And if the person who says these things [about the man standing still and
moving] became even more amusing and was sophisticated enough to say that
whole spinning tops stand still and move at the same time when they have fixed
their center in the same spot and revolve, or that anything else that stays in one
spot while moving in a circle does the same thing, we would not agree, because
such things belong to things that stand still and move not in the same respect of
themselves. Rather, we would say that they have an axis and a circumference in
them and that with respect to the axis they stand still, since they do not incline to
either side, while with respect to the circumference they move in a circle. And
whenever one does incline its axis to the left or the right, or to the front or back,
while it is spinning, we would say that it is in no way standing still. (436d4-e5)

This "sophisticated" counterexample is supposed to prevent Socrates from saying that
some parts of the top move while others stand still. The whole top is supposed to be
moving and at rest (436d5). Presumably, the whole top is moving insofar as it is
revolving and at rest insofar as it does not incline now in this direction and now in that
but continues to occupy the same volume of space (since the axis remains fixed).
Socrates accordingly explains how this top does not violate the principle of non-
opposition by insisting that the top moves in one respect and stays at rest in another (cf.
Laws 893c). So understood, difference in respects is not the same as difference in
parts, and difference in parts is not one of the excusing conditions of the principle of
non-opposition.

---

Although this makes trouble for the standard reading, it does not establish that Socrates has ruled out the possibility that the soul is a complex unity. In fact, he goes on to make room for something very close to the standard reading.

To divide reason from appetite, Socrates restates the principle of non-opposition: Therefore, if sometime something draws the soul back when it is thirsting, wouldn't that be something in it that differs from the thing that thirsts and drives it like a beast to drink? For, as we say, the same thing would not do opposites [1] by the same thing of itself [τῷ αὐτῷ ἑαυτοῦ] [2] about the same thing [περὶ τὸ αὐτὸ] [3] at the same time [ὢμα]. (439b3-6)

This restatement differs from the original in three ways. The first is trivial: this specification leaves out the possibility of "undergoing [πάσχειν]" opposites. The second is potentially significant, but intelligible: "in relation to the same thing [πρὸς τὰ ὑπὸν]" has become "about the same thing [περὶ τὸ αὐτὸ]," perhaps because Socrates assumes that mental experiences are related to other things specifically because they are about other things. (Mental experiences, we might say, are characterized by intentionality.) Third, in place of "in the same respect [κατὰ τὰ ὑπόν]" Socrates has now said "by the same thing of itself [τῷ αὐτῷ ἑαυτοῦ]." This is of some consequence.

Recall the human who stands still while moving his arms and head: "one thing of him is standing still and another [thing of him] is moving" (436d1). According to the trouble-makers, because Socrates does not say that the human moves in one respect and stands still in another, the case does not satisfy any of the three excusing conditions of the principle of non-opposition. It is not an allowable case of one thing being a subject of opposites but is instead a case of a plural subject. But this case does satisfy one of the three excusing conditions of the restated principle of non-opposition. The restated principle allows a human to move "by one thing of him" and stand still "by another thing of him."

Why has Socrates restated the principle to such effect? Perhaps the restated principle is just a more specific version of the original principle, tailored to the case of psychological conflict. On this view, something close to the standard reading was right all along: to φ with one part of oneself while ψing with another is one way of ψing in one
respect and ψing in another. (It is not the only way because the case of the top 
highlights another, though it might be the only way relevant to psychological conflict.)
So on this view, the trouble-makers misunderstood Socrates' characterization of the 
human who stands still and moves. They took him to be saying only that one thing of 
him stands still and another moves, whereas he is also saying that the human moves in 
one respect and stands still in another.

Or perhaps the trouble-makers were right about the original principle of non-
opposition and the human who stands still and moves, but they missed the way in which 
the restated principle of non-opposition blocks the trouble-making implication. On this 
view, the original principle emphasizes the plurality required of an apparent subject of 
simultaneous opposites: we should not say, strictly speaking, that the human moves in 
one respect and stands still in another. But Plato does not mean to imply that the 
human body is not a whole comprising parts some of which can move while others are 
at rest. The restated principle makes this explicit, since it allows us to say that the 
human moves by one part and stands still by another. By putting the principle in both 
ways, Socrates emphasizes both the plurality and the unity of the human body and, by 
implication, the human soul.

---

21 For an attempt to defend the standard reading more thoroughly by rejecting the Stalley-
Bobonich-Lorenz account of the spinning top, see Price 2009.

22 In support of this, one might make two points (cf. Price 2009). First, one might point to the 
question Socrates asks just before he raises the first potential counter-example (436c6-7): "Can 
the same thing stand still and move at the same time in the same respect?" If this question 
focuses Glaucon's attention on ways in which a thing can simultaneously stand still and move in 
different respects, the two potential counter-examples could highlight two different ways in 
which a thing can stand still in one respect and move in another. Second, one might question the 
trouble-makers' construal of 436d1. It is possible to take the τὸ μέν and τὸ δὲ to be accusatives of 
respect and not nominatives, in which case Socrates reports that "we should say... rather that he 
stands still in one respect and moves in another." To reject this reading, Lorenz (2006, 23n9) 
appeals to the parallel case of the archer at 439b8-c1, where Socrates says that one hand moves 
one way and the other another, and not that archer moves one way with respect to one hand and 
another way with respect to the other. But the archer enters after Socrates has restated the 
principle of non-opposition and traded the more general "in the same respect" clause for the more 
specific "by the same thing of itself" clause. That would explain the general talk of respects for 
the human standing still and moving and the more specific talk of parts for archer.

23 Adam (1905, ad 439b) allows only that by the restatement "Plato betrays a sense of the unity of 
the soul." Bobonich (2002, 232-233 with 530n22) goes so far as to deny that restatement differs 
significantly, by appealing unpersuasively to the position of γε in 439b5. But Lorenz (2006, 27) 
appreciates the passage's significance.
Whether the restated principle clarifies or substantially corrects the original, it explicitly allows the soul to be a single thing comprising multiple parts and not merely a many. Thus, when Socrates answers his original question—does a person, a soul, do diverse activities by diverse elements of itself, or does she do each by the whole soul?—he stresses that the diverse sources of an agent's actions are elements in or of a single soul. He says that sometimes when "there are thirsty people who are unwilling to drink … there is something in their soul bidding them to drink, and there is something in their soul, forbidding them to do so, something that is different and that overrides the thing that bids" (439c3-8). He quickly labels "that by which the soul calculates the calculating thing of the soul and that by which it lusts, hungers, thirsts, and gets excited by other appetites the uncalculating and appetitive thing, companion of certain indulgences and pleasures" (439d4-8, translating d5-8). Socrates takes himself to have shown not merely that there are multiple psychological sources within us but that "these two kinds are distinguished in the soul" (439e1-2).

This leaves our puzzle. Although Socrates has left room for the soul to be a complex whole comprising multiple parts, he has also made it clear that the parts are separate sources of psychological activity (appetite bids, reason forbids). The rest of the Republic only makes this clearer, as each of the separate sources of psychological activity

---

24 Among others. I set aside here the problems concerning how many parts Socrates' argument establishes. One of these problems concerns whether he is entitled to think that there is just one appetitive part: the heterogeneity of its desires suggests that there can be conflicting appetitive desires. (Compare the strategies offered by Reeve 1988, 124-131; Irwin 1995, 203-217; Price 1995, 46-48; and Lorenz 2006, 13-52.) Another concerns whether he is entitled to think that there is a spirited part in addition to the appetitive and rational parts. (See Cooper 1984.) A third concerns whether there might not be some part(s) beyond reason, spirit, and appetite. (See 443e and Kamtekar 2008.) My problem, concerning how the parts constitute a whole soul, applies however many parts there are and however they might be subdivided.
activity is characterized as agent-like. What, then, explains how these separate sources are parts of a single soul?

3. Failed explanations of the soul's unity

One might be tempted to say that nothing psychological explains the unity of a soul. Only the brute fact of embodiment renders my rational, spirited, and appetitive elements parts of me and leaves your psychological elements out. One might offer two reasons for this suggestion. First, the soul is supposed to explain the distinction between a living and dead body (353d9-10). The explananda are the vital activities of a living body. That body's soul is the explanans. But if we are dealing with just one living body, we should be dealing with just one soul to explain it. Second, if the soul by itself is a simple unity whose embodied complexity is due to embodiment (611a-612a with 518d-e), the embodied soul's complex unity might be best explained not by the soul

---

25 This has been widely recognized, at least since Moline 1978, although there are debates about exactly what characteristics should be imputed to the different parts, and especially appetite (cf. Bobonich 2002, Lorenz 2006, and Moss 2008). Anyone who takes the parts of the soul to be merely conceptual must somehow explain away Socrates' descriptions of the soul-parts as distinct agent-like subjects of psychological states and activities. But there are good reasons to take these seriously (see Kamtekar 2006). Four reasons to deflate them are unconvincing. First, Price (2009, 10) argues that if the parts are distinct subjects, Plato cannot account for the unity of consciousness (what Price calls "co-consciousness"). But this is prematurely pessimistic. Next, Shields (2001) argues that if soul-parts were more than conceptual, then the soul could not be essentially simple and thus could not be immortal. But in Republic X (611b-612a), Plato entertains the possibility of a complex soul being immortal and the possibility of the same soul being simple when disembodied and being necessarily complex when embodied (see also n. 26). Shields (2001, 147-148) also argues that because the principle of non-opposition can establish a distinction among conceptual parts, it cannot establish organic or aggregative parts. But this does not follow. The principle establishes multiplicity without characterizing the multiple entities or their relation to each other. Finally, Shields (2001, 148-151) argues that Socrates does not establish that the soul is essentially a complex unity since the true nature of the soul might not be multiform (611b-612a) and since the Book Four argument in any case uses a contingent a posteriori premise ("Our souls sometimes both have an appetite to drink and refuse to drink" [439c1-2]). But Socrates does not need to show that the soul is essentially multiform; he can and does allow that the disembodied soul might be simple. He does mean to show that the embodied soul is necessarily multiform—notice the definitions of the virtues in terms of the distinct parts of the soul, though virtues are found in people free of internal conflict—but the a posteriori premise can and should be construed as the existence proof for the modal premise Socrates needs, which is that every embodied soul can experience conflict (cf. Bobonich 2002, 235).
itself, since the soul explains simple, disembodied unity, but by precisely what makes it complex, the body.

These reasons are not convincing. The first gives at best a reason why we should want an account of the soul's unity; it does not actually give that account. The second is perhaps unduly speculative, since Socrates in Book Ten pulls up short of saying that the disembodied soul is partless (612a3-6). But more importantly, the basic claim that the soul's unity is due to the body flouts two basic and widely attested Platonic commitments. First, it flouts the Platonic dictum that the soul rules the body. It also makes the soul's unearned unity an extrinsic property, dependent upon something external to the soul, though Plato's dialogues invariably treat unity as an intrinsic property. In the Republic, Socrates says that the just person puts himself in order, is his own friend, and harmonizes the three parts of himself like three limiting notes in a musical scale—high, low, and middle. He binds together those parts and any others there may be in between, and becomes entirely one thing out of many, moderate and harmonious. (443d5-e2)

This unity is internal harmony, and by calling the unified soul moderate, Socrates means that psychic unity, just like the unity of the city (cf. 462b-e), is defined by internal agreement (442c9-d2, cf. 432a6-b1). This is the earned unity of the virtuous few, and not the unearned unity we are looking for. But this specification nevertheless casts doubt on taking unearned unity to be an extrinsic property.

At this point, one might try a different tack. If the human soul contains three distinct sources of psychological states and activities but is nevertheless a morally

---

26 Before and after 611a-612a in Book Ten, he gives no support to the idea: his argument for immortality assumes nothing about the simplicity of the soul, unlike the so-called "Affinity Argument" in Phaedo 78b-84b (see Brown 1997, and contrast Shields 2001, 143-144) and some of what Er says is more readily intelligible if disembodied souls retain spirited and appetitive elements in them. Moreover, for what it is worth, the Timaeus and Phaedrus suggest divergent answers to the question of whether the disembodied soul is complex or simple. So no matter how suggestive 611a-612a might be, it is difficult to be confident that Plato held a settled view on the question.


28 I here concentrate on the Republic and in the next section draw help from the Phaedrus and Gorgias. For evidence in the Theaetetus, Sophist, and Philebus, see Harte 2002.
responsible agent, perhaps it is quite literally a "plural subject" or corporate agent. On this view, the psychological unity that belongs to a human being is nothing more and nothing less than what belongs to a group such as a Parent-Teacher Organization, and the practice of holding a human being morally responsible is akin to recognizing collective responsibility.

This approach needs an account of collective agency. What makes a collection a plural subject and thus a single agent? One popular answer is that multiple agents become a plural subject by sharing an intention. This can work for all sorts of plural subjects, and it can even explain how a plural subject becomes a kind of unity. Moreover, it would seem to fit well the Republic's insistence that earned unity is a matter of agreement among the parts of the soul. But there is some reason for doubt. Given the different aims that Socrates attributes to the three parts of the soul, it seems plausible that some vicious people will have psychological elements that fail to share intentions. What would explain the unity of such a soul, and thus how it is the locus of moral responsibility? Well, perhaps a group can be a plural subject by sharing an intention that its members do not all recognize, because the relationships within the group give all its members intentions whether they recognize them or not. Perhaps the psychological elements in even a deeply conflicted soul share intentions in this way.

This resembles Plato's account in some important ways. The unearned unity of the soul is, like earned unity, a function of agreement among its parts, but unearned unity does not require each part of the soul to recognize the aims that it shares with the other parts. So the moral responsibility of the whole soul is akin to the collective responsibility of a corporate agent. But Plato does not take the soul to be merely a corporate agent. For one thing, at least some corporate agents are formed and dissolved voluntarily, by convention, whereas the human soul is a natural composite (cf. 588d5-6). There are natural relations among the soul's parts that secure the unity of the soul. This is a good thing, since Plato needs more than a plural subject if his account is

---

30 See, e.g., Velleman 1997.
supposed even potentially to explain the coordinated source of all vital activities and especially the unity of consciousness. To grasp Plato's account of natural unity, we need a fresh start.

4. A fresh approach to unity

In the *Phaedrus*, Socrates tries to lead young Phaedrus to take up the genuine art of oratory and to reject what ordinarily passes as that art. Among other things, he claims that ordinary teachers of oratory give insufficient attention to how an expert speaker fashions his speech into a whole (269c1-5). To motivate this claim, he appeals to the difficulty of crafting a tragedy, and Phaedrus readily sees that a tragedy must be a "system" whose elements fit with one another and with the whole (268d3-5). Socrates says, "Every speech, since it has a body of its own, must be composed like a living being, so that it is neither headless nor footless but instead has things in the middle and at the extremities which are written to fit each other and the whole" (264c2-5). On this account, something is made a whole by the special arrangement of its parts.

Although the *Phaedrus* applies this account only to speeches, tragedies, and living bodies, the *Gorgias* encourages further generalization. There Socrates maintains that every craft, and not just tragedy and oratory, confers special organization to its products (503d-504e, cf. 506c-508a), and he suggests that the same kind of order or organization is manifest in the natural world (507e-508a). In the *Gorgias*, he does not talk of unity or wholeness exactly, but the language he does use (of order [κόσμος] and organization [τάξις]) perfectly fits what he says in the *Phaedrus* about what makes elements into parts of a whole. Encouraged by the *Gorgias*, I attribute to Plato this perfectly general account of complex unities (wholes): diverse elements constitute a

---

32 I cite the text in Burnet 1901, and translations are mine.

33 One might want to insist, more precisely, that something is made a whole by the arrangement of things that are, because of the arrangement, its parts. I forgo such precision here.

34 Text: Dodds 1959.

35 As Dodds (1959, *ad* 504a1) sees.
complex unity, a whole, if and only if they are suitably organized in relation to each other and in relation to the whole that they constitute.

But what makes diverse elements suitably organized in relation to each other and in relation to the whole that they constitute, as opposed to being organized in some unsuitable way? In the *Phaedrus*, Socrates maintains that if the elements of a speech are not organized well, then the speech not only fails to constitute a whole, its speaker fails "to speak these elements persuasively" (269c2-3). That is to say, since oratory is the art of leading souls through speech (261a7-8, cf. 271c10), a speech without good organization fails to perform its function well. This explanation also generalizes well to other artifacts and living bodies. A shipbuilder's product, if it is not a whole of suitably ordered parts, will fail to be seaworthy. A living body, if it is not a whole of suitably ordered parts, will fail to be able to perform its vital activities.

Plainly, there are degrees of organization and degrees of success here. A speech (or ship or living body) need not be perfectly organized to be persuasive (seaworthy, vitally active), but the best speeches (ships, living bodies) will be more persuasive (seaworthy, good at vital activities) by displaying greater organization than lesser ones. Still, in every case, the range of permissible organization has a lower limit. Below this limit, there is nothing performing the function of the relevant kind, and so nothing of the relevant kind at all: sometimes we have a heap of potential ship-parts but not a ship, a string of words but not a speech, or an assortment of limbs but not a living body.

In sum, then, the *Phaedrus*, with help from the *Gorgias*, suggests that diverse elements are parts of a whole if and only if they are organized in relation to each other and to the whole in such a way that the whole is able to perform its function. Unity is a function of organization, because a complex thing's ability to do what it does depends upon its parts' ability to do what they do, which in turn depends upon their orderly
relations (since no part could do what it does as a part independent of the other parts and the whole). \(^{36}\)

5. Explaining the soul's unity

By now, my hypothesis should be obvious: I suggest that Plato treats the human soul like a speech, a ship, or a living body. The soul's unity is a function of the order or harmony of its parts. The parts' causal relations with each other produce this order or harmony: the more the parts do what they are supposed to do, as parts, the more the soul enjoys unity. On this view, Socrates can and does say that the soul's unity takes the specific form of agreement, since agreement is the harmony or order of minded things. And on this view, the soul can perform vital activities—that is, its function of living—with only some minimal agreement and order, since it requires perfect agreement and order only for excellent or virtuous activities. (That is the distinction between unearned and earned unity.)

In the *Republic*, Socrates does not make this account fully explicit. He does explicitly characterize the just person's maximal unity as a kind of harmony and agreement (462d5-e2, quoted and discussed above), and with his account of justice, he makes clear that the parts produce this harmony by doing what they are supposed to do. But he only hints at how to extend this account of unity to less than perfectly just souls. First, he explains to Glaucon,

> For the sake of a model we were seeking both what sort of thing justice itself is and what sort of man the completely just man would be if he came into being,

\(^{36}\) Harte (2002) finds in Plato's *Sophist, Timaeus*, and (especially) *Philebus* a broadly similar account of wholes in terms of the structure or harmony of parts. The account she finds, however, does not welcome degrees of structure and wholeness or unity. If the *Republic* offered the same account (and see §6 for this), earned unity would be the only kind of unity that souls could enjoy, and in fact, any talk of the whole soul would be merely an unjustified manner of speaking except when we were speaking of the virtuous. But Plato has good reasons for wanting the soul to have unearned unity (see §2 above) and Socrates speaks as though it does (see §1 and §3). Moreover, what Socrates says in the *Republic* about the soul's order and functioning fits the general account of the *Phaedrus* and *Gorgias*, according to which unity comes in degrees. If my account is right, then Harte's is wrong; Plato changed his mind after writing the *Republic*; or Plato has two distinct accounts, one (admitting degrees) for things subject to becoming and another (not admitting degrees) for abstract objects.
and likewise injustice and the most unjust man, in order that, by looking at them and how they seemed to us concerning happiness and its opposite, we would have to agree about ourselves as well, that the one who was most like them would have a portion of happiness most like theirs. (*Rep.* 472c4-d1)

Here Socrates says that the imperfect soul lives more or less well (has a greater or lesser "portion of happiness") according to how closely its order approximates the just soul's. He elaborates in Book Nine, arguing that the perfectly just soul lives best (is most happy) and other souls live more or less well (are more or less happy) according to how closely their order approximates the just soul's (580a-c). Socrates assumes both that there is a correlation between psychological order (that is, justice) and psychological functioning (that is, living) and that each of these features comes in degrees.

These assumptions do not commit Socrates to the thought that imperfect souls enjoy unearned unity. But he also suggests in Book One that a soul needs some minimal justice if it is to perform its function. After arguing that a city or a band of thieves would be unable to achieve an unjust purpose if its members were unjust toward each other (351d), he generalizes the point thus: because injustice sows internal disharmony, it threatens to render "a city, a family, an army, or anything else," including a soul (cf. 352a6-7), "unable to do anything with itself" (351e10-352a3).

Moreover, Socrates offers two causal principles about the interaction of the soul's parts that illuminate the structural order that makes the three psychological elements parts of a whole. The first can be called Plato's hydraulic principle of psychology. Socrates says, "When someone's desires incline strongly for one thing, they are thereby weakened for others, just like a stream that has been partly diverted into another channel" (485d6-8; cf. 328d3-5). This plainly applies to different soul-parts within the same soul. If my spirited desires grow stronger, my rational and appetitive ones grow weaker. Just as plainly, the principle cannot apply to psychological elements in different people: if your spirited desires grow, nothing automatically follows for my rational and appetitive desires. (Of course, you might influence me, and if my spirited desires were to strengthen, too, then my rational and appetitive desires would weaken. But this is a rather different story.)
The second I will call Plato's principle of psychological hegemony. In Books Eight and Nine, Socrates maintains that different kinds of people are ruled by different soul-parts. As he explains this, to be ruled by a soul-part is to take the ends of that soul-part to be one's ends, generally. If I am ruled by spirited desire, I take my good to be honor or victory, the ends of the spirited part. If I am ruled by appetitive desire, I take my good to be bodily pleasure or, perhaps, if I calculate a bit, the best means to achieve bodily pleasure, which is money. This pattern requires certain causal interactions within the soul. Because the spirited or appetitive part of the soul gets more powerful, the rational part, which alone calculates what is good for the whole soul (441e3-4, 442c4-7), begins to calculate that it would be good to achieve what the powerful part wants. When a young man is influenced by his father to pursue wisdom and by everyone else to pursue money and publicly contested honors, he calculates that his good lies in the middle way, pursuing the spirited part's end of honor (549c-550b). When another young man sees his honor-loving father killed or reduced to poverty, he calculates that it would be safer to pursue money, the end of the appetitive part (553a-d).

This principle explains what it is for a soul to be ruled by some part of the soul by explaining what it is for the rational part of the soul to be ruled. Corollaries to the principle explain what it is for the spirited or appetitive part to be ruled. When spirit is ruled by either reason or appetite, the spirited part finds honor in the ends of the ruling part. When appetite is successfully ruled by either reason or spirit, it limits its pleasures to those deemed good by ruling reason or honorable by ruling spirit (and thus good by ruled reason).

Again, the principle of psychological hegemony and its corollaries explain direct connections in me without imputing direct connections between your spirited element and me. That your spirited element has grown so strong as to dominate your soul does not force my rational part to determine that honor or victory are my good (or yours). Again, you might influence me. If my spirited desires were to strengthen, too, then I

---

37 The evidence for this principle is diffuse; it emerges as the best explanation of what Socrates says in Books Eight and Nine about what it is for one part of the soul to rule.
would calculate that honor or victory is good for me (cf. 549c-550b). But that is not the
direct causal influence that my soul-parts enjoy over each other.

The two causal principles are related. For instance, a rational part can come to
accept that honor or victory is what is good only if it is too weak to grasp what really is
good, but it will be too weak if the spirit has taken much of the soul's "hydraulic power"
away from reason.

Also, these are not the only causal principles Plato would invoke to explain
psychological behavior. Among others, Plato clearly believes that the appetitive part
has its own internal tendency to grow stronger, which requires that it be checked at
every turn (442a). Socrates paints this principle into his imaginative portrait of the soul
at the end of Book Nine, when he says that the just person must continually prevent the
many-headed beast of appetite from sprouting savage heads (589a-b; cf. 571b-572b).
He also assumes it when he asserts that gold and silver need to be kept away from
guardians (416e-417a) and when he predicts that a person with a timocratic soul will
become progressively more money-loving over the course of his lifetime (549a-b). So
there are causal tendencies in the soul that encourage it to become less orderly and
unified.

Still, the hydraulic principle of psychology and the principle of psychological
hegemony explain how the soul tends to organize and unify itself, even when its
appetite grows. Even in these circumstances, there is a tendency for the rational part to
agree with appetite about what is good and a tendency for the spirited part to find some
honor in what appetite pursues. This tendency is a product of the causal relations
among my soul-parts and does not extend to yours, which captures the sense in which
my soul-parts are parts of one psychological system, one soul. They are parts of one
soul because they are parts of an orderly causal structure that manifests itself in every
soul as a tendency toward agreement, a tendency reaches perfect fulfillment in only
some souls. The minimal structure is required for living, its perfect fulfillment for living
well. But not even the most corrupt dissolves to the point of inactivity.³⁸

³⁸ Cf. 608c-611a with Brown 1997.
This account of psychic unity shares some features with accounts of collective agency. Just as some human beings might share aims due to the way they relate to each other in a group—even if they do not realize that they share these aims, or converge on some mistaken view about them—so, too, the soul's parts share aims though they do not realize it or converge on a mistaken view about them. This parallel helps to explain how the whole soul can be held responsible for whatever any part of it does. When one part of the soul does something, the others are causally implicated in a structure that fosters agreement. At the least, each of the others could have been a more powerful obstacle. So it would not be inappropriate to hold the rational part alone responsible in the afterlife for what the other parts had done during embodied life.

Unlike collective agency, though, psychic unity on this account depends upon a purely natural structure of causal relations (cf. 588d5-6). Given the way the soul and body naturally are, the embodied soul comprises parts that causally relate to each other in ways that tend toward agreement. These parts are distinct subjects, but the account presupposes that they are not distinct in exactly the same way that two human beings are distinct subjects. It presupposes that soul-parts naturally share information intimately, though it does not explain exactly how or how much information is shared. It also presupposes that at least much of the shared information manifests itself to the soul as a single consciousness, and again, we might ask questions about exactly how this is supposed to work. The Republic does not address those questions, but it does not render them unanswerable, either.39

6. Sustaining this explanation of the soul's unity

This account of the soul's unearned unity is incomplete. Most obviously, it lacks a precise notion of internal agreement. This calls for some reckoning of the mental content of the various soul-parts, to see how there could be shared content. This would give additional flesh to the account of psychic unity and perhaps explain the unity of

39 Related questions face current philosophers who suppose that modularity is essential to explaining the mind-brain (e.g., Fodor 1983).
consciousness. But for now, the skeleton of the account needs further defense, and I close by addressing suggestions that some souls lack unity.

The first depends upon the city-soul analogy. Socrates tells Adeimantus,

You are happily innocent if you think that anything other than the kind of city we are founding deserves to be called a city ... We need a bigger category to address the others because each of them is a great many cities, not a city ... At any rate, each of them consists of two cities at war with one another, that of the poor and that of the rich, and each of these contains a great many. (Rep. 422e1-423a2)

Now, if none but the ideal city is really a city, because every other claimant fails to be a unity, then the analogy of city and soul suggests that none but the ideal soul is really a soul, because every other claimant fails to be a unity. 40

But the analogy of city and soul can only suggest; it cannot prove. There is an open question about which features of cities have analogues in souls and vice versa, and so it is easy enough to suppose that this particular feature of cities does not have an analogue in souls. Moreover, we can justify this supposition. First, cities do not seem to have the same causal regularities governing the interaction of their parts that souls do. In particular, the principle of psychological hegemony has no analogue for the city: that the oligarchs rule does not tend to make the democrats accept the oligarchs' rule. This is not to say that there is no tendency toward shared values in a city: Plato plainly recognizes the power of a shared culture to shape the citizens' values. 41 But shared values do not necessarily produce agreement about who should rule in the city as it does in the soul. Agreement in values just is agreement about who should rule in

40 One might doubt that Socrates is entirely serious here. In Book One, he urges that even a band of thieves must have some justice to achieve its purpose (351c-d), which reinforces the idea that justice comes in degrees and encourages the thought that even an imperfectly functioning community has enough justice to be a community. I am nevertheless inclined to think that Socrates is seriously reserving 'polis' for the most successful communities (cf. Pol. 293e). Perhaps the ensuing discussion above helps to explain why he does this: he might be worried that attributing goodness to every community undercuts the goodness that he wants to predicate of only the best communities. Alternatively, his reasons might run deeper: perhaps the causal relations among human beings, unlike those among the parts of the soul, do not underwrite unearned unity. I thank Anton Ford and Clerk Shaw for discussion of this point.

the soul. In the city, by contrast, the oligarchs and democrats can agree broadly in their values but disagree about whether the rich alone should rule. Second, competing factions in a city do not share responsibility for "what the city does" as parts of a divided soul do. That is why Socrates insists that when Kallipolis is at war with another Greek city, the Kallipolitans will not ravage the entire enemy city but will target the real enemies in the city who are responsible for the war (471a-b). These differences between cities and souls explain why it is reasonable to say that a divided city is not a whole city but not to say that a divided soul is not a whole soul.

But this might seem to prove too much. In Book Eight, Socrates returns to his point about divided cities, and he extends his analysis to divided souls. Having said an oligarchic city is "not one but two, one of the poor and one of the rich" (551d5), he claims of an oligarchic soul, "Such a person would turn out not to be free of civil strife in himself, and he would be not one but in a way twofold [διπλοῦς τις], but for the most part he would have better desires in control of worse" (554d9-e2). This is the strongest evidence that Socrates is willing to deny unearned unity to a human soul. But it shows, in fact, that Socrates is not willing. In the case of divided cities he says forthrightly that they are two (δύο, 422e6 and 551d5), but of the soul he says only that it is "in a way twofold" (διπλοῦς τις, 551e1). The hedge is significant, and well explained by the fact that even the "in a way twofold" soul has parts that are causally related so as to promote agreement. Socrates says explicitly that such a soul "for the most part would have better desires in control of worse," and presumably, if the better desires were to lose control of the worse, then the soul would tend away from oligarchy and towards democracy by jettisoning the distinction between better and worse appetitive desires and treating all equally. The causal tendencies are enough to establish that the oligarchic soul is a single soul.

Cf. Thucydides' presentation of the Mytilenian debate. The Athenians come to understand that the democrats in Mytilene were not responsible for the revolt, because that revolt was led by the oligarchic faction and stopped by the democrats themselves (esp. 3.47.3). It is true that the Athenians come to see this belatedly (3.36) and only by a slight majority (3.49.1). But this is not because the point is reasonably contested: at least, Thucydides suggests that passionate rage has blinded those who want to punish the Mytilenian democrats (3.36.2, 3.44.4).
Socrates' point in this passage is not that the oligarchic person has more than one soul. His point is that the oligarchic person's soul falls far short of the unity that virtue earns. He makes this sort of point often in the Republic, as he is keen to convince his interlocutors that people who do not pursue wisdom suffer psychologically in ways that philosophers do not. But this point deprecates the earned unity of non-philosophical souls, not their unearned unity. Even the most divided soul is still a single soul whose unity is explicable in terms of a structure of parts causally related so as to tend to agreement.

---

43 Recall 608c-611a with Brown 1997. One might say that Socrates' point deprecates their "ethical" unity, not their "metaphysical" unity, but this is misleading. There is just one kind of unity. Every embodied soul possesses it at least to the minimal extent that a complex entity needs to exist, though only some souls possess it maximally.
Bibliography


Research 57: 29-50.