The Form of the Good and the Unhypothetical First Principle in Plato and Aristotle

Hugh H. Benson

(Draft: June 2010)

In recent years Dirk Baltzly and Dominic Bailey have usefully examined the inter-textual similarities between Plato’s discussion of the unhypothetical first principle (τὸ ἐπ’ ἀρχήν ἀνυπόθετον; hereafter UFP)\(^1\) in the (in)famous Divided Line image in the Republic and Aristotle’s ‘defense’ of the law of non-contradiction (hereafter LNC) in Metaphysics Γ. As both of these scholars have helpfully pointed out these two passages contain the only occurrences of ἀνυπόθετον (twice) and its cognates (ἀνυποθέτον (once)) in the entire Platonic and Aristotelian corpus. This feature of these passages alone requires the sort of investigation these scholars provide. Aristotle’s discussion of the UFP in Metaphysics Γ ought to inform our reading of Plato. Nevertheless, I maintain that their respective accounts of Plato’s UFP in light of Metaphysics Γ leave unaddressed salient evidence from the Republic itself. In brief and very roughly, Bailey’s account leaves unaddressed the evidence from the Republic that the UFP is the form of the Good, while Baltzly’s account leaves unaddressed the evidence that the method by which one reaches the UFP according to Plato is the method of hypothesis as introduced and developed in the Meno, Phaedo, and the Republic. I do not suppose that this unaddressed evidence refutes their respective accounts. Indeed, I think it can and should be integrated into their accounts. But when it is, it leaves open the very questions Bailey’s and Baltzly’s accounts sought to answer - roughly what grounds the completion of Platonic inquiry?

The Form of the Good as Plato’s UFP

\(^1\)Throughout I will be frequently using the definite article, contrary to both Bailey and Baltzly, who take Plato to be committed to a plurality of UFPs. I take Plato’s use of “τοῦ παντὸς” in ἵνα μέχρι τοῦ ἀνυποθέτου ἐπὶ τὴν τοῦ παντὸς ἀρχήν ἰόν at 511b6-7 to suggest Plato’s commitment to a unique UFP (or at least to be concerned with a specific UFP here in this passage), but nothing in what follows depends on this. What will be important is that the form of the Good is a UFP, whether or not there are others.
Let me begin by defending what I take to be a governing condition on any account of Plato’s UFP - viz., that Plato’s UFP is identical to the form of the Good.\(^2\) We should remember that the three images - the Sun, the Line, and the Cave\(^3\) - found in Books VI and VII of the Republic are explicitly offered in an attempt to answer Glaucon’s ‘What is the Good?’ question, on the assumption or pretense that Socrates does not know what the Good is. At Republic 502c9-e3 Socrates has completed his argument for the possibility of Kallipolis and turns to the remaining question concerning how the philosopher-rulers necessary for the possibility of Kallipolis will arise in the city. Socrates explains that prospective philosopher-rulers will need to be tested in labors, fears, and pleasures, and exercised in the most important subjects (μέγιστα μαθήματα). When Glaucon asks what Socrates means by ‘the most important subjects’, Socrates distinguishes between a longer and shorter road, and maintains that the philosopher-ruler must take the longer road, the only road that leads to the possession of the most important subject (τοῦ μεγίστου ... μαθήματος; 504d2-3).\(^4\) Glaucon predictably asks what the most important subject is, and Socrates immediately replies that it is the form of the Good (ὅτι γε ἂ τοῦ ἄγαθοῦ ἰδέα μέγιστον μάθημα; 505a2) of which, he avers, we fail to have adequate knowledge. Socrates then goes on to describe a number of features of the Good, including that it is not knowledge or pleasure, that if one fails to know it, one fails to know anything fine and good, and that the soul “divines that it is something, but is perplexed, and unable to adequately grasp what it is or to acquire the sort of stable beliefs it has about other things” [505d11-506a3; based on Grube/Reeve trans.]. Glaucon again predictably asks “whether you consider the good to be knowledge or pleasure or something else altogether” (506b2-4; Grube/Reeve trans.). After professing his ignorance of the Good again, Socrates proposes to abandon describing directly what the Good is, ...

---

\(^2\)It may be surprising that the claim that Plato’s UFP is identical to the Form of the Good should require defense. Certainly many scholars have maintained this identity (see, for example, (Robinson 1953:159-160), (Santas 1980:252–256), and (Denyer 2007:306)), and some without defense. But Bailey has suggested to me in conversation that he wonders about this identity, and (Baltzly 1996:157) explicitly maintains that the reasons offered on behalf of the identity are not decisive. (Bedu-Addo 1978:124), for one, denies the identity.

\(^3\)Throughout I will refer to three images with caps, i.e. as the Sun, the Line, and the Cave.

\(^4\)One is hard pressed I think to explain the change from the plural to the singular. Although see Bailey’s view that there is more than one UFP for Plato.
and instead offers to state what appears to be an offspring of the Good and most like it - the sun. Thus, the image of the Sun is explicitly introduced as an attempt to illuminate the Good. Socrates follows the Sun image, with the image of the Line and the image of the Cave, and when Socrates concludes his description of the Cave at 517a-c (quoted at length below), he reminds Glaucon that all of this (the Sun, the Line and the Cave) has been offered in this attempt to illuminate what the Good is like. Thus, whatever else we want to say about these three images, Socrates explicitly introduces them in an attempt to illuminate the form of the Good.

Moreover, what Plato says about the relationships among the images also indicates they are meant to illuminate the Good. After concluding the Sun image with the well-known passage in which the Good is credited with making the objects of knowledge both knowable and ‘beable’, although the Good is not being but beyond being, Socrates transitions from the Sun to the Line as follows, beginning with Glaucon’s exclamation:

By Apollo, what a daemonic superiority!
It's your own fault; you forced me to tell you my opinion about it.
And I don't want you to stop either. So continue to explain its [the Good’s] similarity to the sun, if you've omitted anything.
I'm certainly omitting a lot.
Well, don't, not even the smallest thing.
I think I'll have to omit a fair bit, but, as far as is possible at the moment, I won't omit anything voluntarily.
Don't. [509c1-11; Grube/Reeve trans.]

It is clear that the Line is intended as a further elaboration of the Sun, and as such as further elaboration of form of the Good. Nevertheless, neither the sun nor the Good are explicitly mentioned in the Line. Something in the Line must correspond to the sun in the Sun and so the Good

The transition between the Line and the Cave, in which the sun and the Good explicitly reappear, is more puzzling, because more abbreviated. All that Plato writes is

Next, I said, compare the effect of education and of the lack of it on our nature to an experience like this: ... [514a1-2; Grube/Reeve trans.]

But later, in concluding the Cave and the images as whole, we are famously told to ‘fit the images’ together as follows:
This whole image, Glaucon, must be fitted together (προσαπτεόν) with what we said before. [A1] The visible realm (τὴν μὲν δὲ ὅψεως φανομένην ἔδραν) should be likened to the prison dwelling, and the light of the fire inside it to the power of the sun. [A2] And if you interpret the upward journey and the study of things above (τὴν δὲ ἀνάβασιν καὶ θέαν τῶν ἄνω) as the upward journey of the soul to the intelligible realm (τὴν εἰς τὸν νοητὸν τόπον τῆς νοητῆς ἄνωθεν), you'll grasp what I hope to convey, since that is what you wanted to hear about. Whether it's true or not, only the god knows (θεὸς δὲ ποιοῦ ὁ θεός οὐδεὶς οὐ θυγατέρει). But this is how I see it: [B] In the knowable realm, the form of the good is the last thing to be seen, and it is reached only with difficulty (εἰς τὸν αὐτὸν καὶ μόνος ῥᾳδιὰ). And once one has seen it, however, one must conclude (συλλογιστεά) that it is the cause of all that is correct and beautiful in anything (ὠς ἄρα πάσι πάντων αὕτη ὁρθὸν τε καὶ καλὸν αἰτία), that it produces both light and its source in the visible realm (ἐν τῷ φῶς καὶ τὸν τούτου κύριον τεκοῦσα), and that in the intelligible realm it controls and provides truth and understanding (ἐν τῷ κυρίω κυρίω ἀληθείαν καὶ γνῶσιν παρασχομένη), so that anyone who is to act sensibly in private or public must see it (ὅτι δὲ ταύτῃ ἰδεῖν τὸν μέλλοντα ἐμφανῶς πράξειν ἡ ἱδία ἡ δήμοσις). [517a8-c5; Grube/Reeve trans.]

While considerable controversy surrounds precisely how these images are to be ‘fitted together’, again it is clear that the Good must be found somewhere in the image of the Line if the images are in fact to be ‘fitted together’. But, again, the form of the Good is not mentioned in the Line. Something in the Line must correspond to the form of the Good.

This last passage, however, may help to identify the form of the Good in the Line. In the [A] portion of the passage Socrates distinguishes between the visible and intelligible realms. This distinction is not explicitly mentioned in the Cave, but is explicit in both in the Sun and the Line. The goal of the [A] portion is to align the Cave with the preceding images. In [A1] Socrates aligns the prison dwelling and the fire in the Cave with the visible realm and the sun in the Sun. Remember the sun is (explicitly) absent from the Line. In [A2], however, Socrates aligns the upward journey inside the cave and the study of the things outside the cave with the upward journey to the intelligible realm in the Line. This is indicated by the fact that the Sun does not refer to a journey or movement of the soul. The upward journey which is aligned with the upward journey in the cave and the study of things outside the cave must refer to the movement

5See, for example (Ferguson 1921), (Joseph 1948:41), and (Robinson 1953:181–190) who argue that the Cave is not meant to correspond (at least closely) to the Line, and (Gould 1955:174–175), (Malcolm 1962:40–41), (Malcolm 1981:62), (Burnyeat 1984:228 n. 38), (Smith 1999:202), (Schofield 2007:230) who maintain that Cave is meant to correspond (at least roughly) to the Line, while (Murphy 1951:156) takes the Line and Cave to correspond to the Sun.
of the soul described in the Lb portion of the Line. It is true that 517b4-5 refers to the upward journey to the intelligible realm, while the only explicit journey of the soul in the Line is within the intelligible realm. But it remains that the only journey of the soul in either image prior to the Cave is the journey of the soul in the top two sections of the Line. Having connected the Cave to both the Sun and the Line, in [B] Socrates says that “[i]n the knowable realm (ἐν τῷ γνωστῷ), the Form of the Good is the last thing to be seen, and it is reached only with difficulty.” The claim that the form of the Good is the last and considerably difficult thing to be seen is familiar from the Cave, but the claim that this process takes place in the intelligible realm comes from the Line given [A2]. But the only candidate in the Line for something that is last thing to be

I will be assuming the following image of the Line and the names of the sections in what follows.

```
<table>
<thead>
<tr>
<th></th>
<th>L4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L3</td>
</tr>
<tr>
<td></td>
<td>L2</td>
</tr>
<tr>
<td></td>
<td>La</td>
</tr>
<tr>
<td></td>
<td>L1</td>
</tr>
</tbody>
</table>
```

See (Robinson 1953:187).

The closest one comes to a movement or journey of the soul in the Sun is at 508c3-d10, but even so the dynamic nature of this image is considerably understated compared to the top two sections of the Line. It is for the reason that I find (Robinson 1953:182)’s view that the “Cave is mostly about motions or progressions from point to point; but the Line is not” difficult to sustain. The top two sections of the Line (L4 and L3) are full of motion words, and most of the Line is devoted to describing these two sections.

The distinction between a visible realm (τῷ ὀρατῷ τόπῳ) and an intelligible realm (τῷ νοητῷ τόπῳ) is never mentioned in the Cave until our current passage. It is mentioned, however, in both the Sun and the Line. Plato’s license with the names of the divisions, and their sub-divisions is notorious. Plato introduces the division in the Sun by distinguishing between visible things (τὰ ... ὀρατά) and intelligible things (τὰς ... ἰδέας νοεσθάν) at 507b9-11. By 508b12-c2 he has identified the visible realm (τῷ ὀρατῷ) and the intelligible realm (τῷ νοητῷ τόπῳ). The Line begins by reminding us of these two places (ὁρατου καὶ νοητου ... τόπου) and the two kinds (διί τε εἴδη, ὀρατόν, νοητόν καὶ τοῖς ὀρατομένοις γένοις καὶ τοῖς νοημένοις) at 509d1-8, but by 510a5-510b1 he is distinguishing between the opinable and knowable (τὸ δοξαστὸν ... τὸ γνωστὸν) and then immediately returns to the intelligible realm (ην τοῦ νοητοῦ τόμη) at 510b2-3 and again at 511a3 (νοητοὶ) and 511b3 (τοῖς νοητοῖς). At 533e4-534b2 the two realms or sections are designated as opinable and intelligible (δοξαστὸν τε καὶ νοητοὶ) and in our present passage the intelligible realm is referred to as both intelligible (τῶν νοητῶν τόπων) and knowable (τῶ γνωστῶ). The
discovered and only after great difficulty is the UFP. In the knowable division of the Line the
last thing the soul comes to in its investigation from hypotheses is the UFP before reversing itself
and reaching a conclusion. And while Socrates does not explicitly refer to the difficulty of the
examination from hypotheses that culminates in the UFP, perhaps the distinction between the use
(or misuse) of sense experience in L3 and the use “only of forms themselves, moving on from
forms to forms, and ending in forms” is meant in part to capture this difficulty. Having
suggested the identity the form of the Good in the Cave with the UFP in the Line in [B], Socrates
goes on in [C] to explain the nature of this UFP based on the Sun.

All of this, then, presents a strong prima facie case for the identity of the form of the
Good and the UFP. The stage directions surrounding the three images indicate that the images
are each intended as progressively more revealing illuminations of the form of the Good.
Nevertheless the Line makes no explicit mention of the form of the Good, and one is hard
pressed to find anything else in the Line other than the UFP that can plausibly be thought to
correspond to it. Moreover, the concluding passage in which Socrates explains, to the extent that
he does, how the images are to be ‘fitted together’ reinforces this correspondence.

Consequently, while I do not mean to suggest that denying the identity of the Plato’s UFP and the

cognitive state associated with L4 is designated as intelligence (νοῦν and νοητική) at 511d6-e5 and 511c3-d5, but as
knowledge (ἐπιστήμη) at 533e4-534b2.

10For a defense of how we are to understand the nature of the procedure described in L4 and the distinction just
mentioned see (Benson forthcoming).

11An additional consideration in favor of this identity derives from Bailey’s plausible contention that the priority
condition is a sufficient condition of Plato’s UFP; see pp. xxx below. Plato is explicit in the run up to the images
that knowledge of the form of the Good is prior to knowledge of anything fine and good; see 505b2-3. And the
claim in the Sun that the Good is the cause of all knowledge (see 508e3-4) may be an indication that Plato would
endorse the stronger thesis that knowledge of the Good is prior to knowledge of anything at all. If so, then, Plato’s
form of the Good satisfies a sufficient condition for being the/a UFP.

12See (Baltzly 1996:164–165 n. 34) who provides (although does not enthusiastically endorse) the following
arguments for identifying the unhypothetical first principle with the Good: “The first argument is an analogical one:
the Sun (Good) plays a unique epistemic role in Plato’s analogy of the Sun. The unhypothetical first principle in the
Line seems to play a unique role too. So it concludes, probably, these things are one and the same. The second
argument is similar: the unhypothetical starting point is the end of the upward path in the Line. The vision of the
Sun is the final stage of the prisoner’s progress in the figure of the cave. So, the starting point of the Line is the
Good. Finally, Plato calls the Good the end of the intelligible ... at 532b. The intelligible, of course, is his term for
the upper half of the doubly divided line. If we assume that the unhypothetical starting point of the Line is uniquely
situated within the order of the intelligibles, then it might well be this very end point. The apprehension of the Good
is, of course, the final phase of dialectic in the image of the Cave.”

6
form of the Good is beyond the pale, I do maintain that this identity ought to be a governing assumption of any account of Plato’s UFP. Compelling evidence, indeed, should be required for denying the identity of the Plato’s form of the Good and his unhypothetical first principle (UFP).

Bailey’s Account: The Form of the Good is not Incorrigible

Given this identity, let us now turn to Bailey’s account of Plato’s UFP. Bailey argues that in *Metaphysics* Γ.3 Aristotle attributes the following three features to unhypothetical principles:

1. Unhypothetical principles are such that error about them is impossible [βεβαιότατη δ’ ἄρχη πασῶν περὶ ἴν διψυχεσθήναι ἀδύνατον] (1005b11-12). (I shall call this the incorrigibility condition.)

2. Unhypothetical principles are necessarily the most intelligible principles [γνωριμοτάτην τε γὰρ ἀναγκαῖον εἶναι τὴν τοιαύτην] (1005b13). (I shall call this the intelligibility condition.)

3. Unhypothetical principles are necessarily part of the equipment of anyone who grasps any of the things that are [ἡν γὰρ ἀναγκαῖον ἐχεῖν τῶν ἄλλων ἐνέπληθα τῶν ὀντῶν, τοῦτο ὅτι ύπόθεσις] (1005b15-16). (I shall call this the priority condition.)” (Bailey 2006:102–103)

Bailey goes on to argue that these same conditions are features of Plato’s UFP as well. His argument that Plato’s UFP meets the first two conditions goes roughly as follows.

Plato’s method of hypothesis proceeds by proposing a hypothesis which will help determine an answer to the original question whose answer one seeks to know. Then one checks to see whether the consequences of that hypothesis cohere with one another (whatever precisely that is taken to mean). If the consequences do not cohere with one another, the hypothesis is false, and according to Bailey one needs to start over. If, however, the

---

13Bailey thinks that the wording of the priority condition in Γ indicates that it is a sufficient but not necessary condition for being unhypothetical, unlike the first two conditions which are necessary and sufficient; (Bailey 2006:103). I take it he has in mind the following ἐν γὰρ ἀναγκαῖον ἐχεῖν τῶν ἄλλων ἐνέπληθα τῶν ὀντῶν, τοῦτο ὅτι ύπόθεσις. I suspect that this may be too fine a reading of the text in any case whether Plato took this latter condition as merely sufficient and not necessary turns on whether Plato allows a plurality of UFP’s which in an earlier note I proposed to set aside; see n. 1 above and (Bailey 2006:106). I found Bailey’s argument that Plato is committed to this third condition (whether as merely sufficient or both necessary and sufficient) quite persuasive; see n. 11 above.

14I put it in this rather vague way because Bailey does not take the falsity of the hypothesis to determine an answer to the original question, despite what I take to be the rather strong evidence from the *Meno* that Plato takes (at least the original) hypothesis to be truth-functionally equivalent to the answer to the original question. But again this need not detain us.

15Bailey offers his own account of this procedure in (Bailey 2005).
consequences do cohere with one another, then one must check to see whether this hypothesis is a member of a set of consequences of another higher, perhaps more general hypothesis, and see whether those consequences cohere with one another. According to Bailey, the method continues - checking for coherent consequences and checking for higher hypotheses - until one reaches a hypothesis one does not need to check any further. This occurs when one reaches a hypothesis that meets either the incorrigibility condition or the intelligibility condition. When one reaches a hypothesis about which one realizes that error is impossible, one can cease checking its truth. Alternatively, when one reaches a hypothesis that is as intelligible as a hypothesis can be, i.e., the most intelligible, one can cease looking to explain it. The key for Bailey is that unless a hypothesis meets the incorrigibility or intelligibility condition we have no rational grounds for bringing this process to a successful conclusion. Consequently, given the nature of Plato’s method of hypothesis, charity requires that, like Aristotle, Plato too took his UFP to meet the incorrigibility and intelligibility conditions, if his method of hypothesis is to have any chance of success.

Now I am not certain what Bailey intends by the incorrigibility condition, but I want to maintain that the identity of the the form of the Good with Plato’s UFP excludes one reasonably

---

16See n. 17 below.

17I confess I am not certain whether Bailey thinks that the incorrigibility and intelligibility conditions are each individually sufficient and their disjunction necessary for a Platonic UFP, or whether they are jointly necessary and sufficient. Bailey writes that “The incorrigibility and intelligibility conditions appeared to be both necessary and sufficient for being unhypothetical” which may suggest the latter. The charity argument, which to be fair is not explicitly in Bailey’s text (although it is suggested at (Bailey 2006:102–103)), seems to require the former. I do not see why the intelligibility condition is not alone sufficient, since intelligibility entails truth (as Aristotle may be indicating in περὶ γὰρ ᾧ μὴ γνωρίζουσιν ἄπαθῶται πάντες). Incorrigibility may be alone sufficient as well, especially if it is recognizable. But it does not matter for our present purposes, for in either case I will be arguing that Plato supposes that his UFP requires further checking by means of the method of hypothesis.

18In fact, Bailey says very little about the nature of this incorrigibility condition. Nevertheless, I have been struck by how difficult it has been to get a clear statement of the nature of incorrigibility. See, for example, Kind (http://www.iep.utm.edu/introspe/) and (Alston 2010:433). This is especially true in light of the notorious difficulties surrounding Plato’s commitment to propositions, and propositional knowledge, to which I will turn in a moment. Consequently, I am less concerned to suggest that Bailey is subject to the worries that follow, than to maintain that the identity of the UFP with the form of the Good excludes understanding incorrigibility (or intelligibility for that matter, cf. n. 19 below) as a condition of the UFP that independently enables one to recognize its truth. If incorrigibility is not understood in this way, I am uncertain what work the UFP’s incorrigibility does in the process of inquiry beyond the assertion that the UFP is true.
plausible understanding of this condition. The charity argument suggests that the role of the incorrigibility condition is to provide a feature of the UFP that enables the practitioner of the method of hypothesis to know when to stop checking one’s hypotheses for truth. The incorrigibility of the UFP guarantees its truth. And recognizing the incorrigibility of a proposed hypothesis provides sufficient grounds for concluding the inquiry, for in recognizing its incorrigibility one recognizes one cannot be mistaken about it. Given the identity of Plato’s UFP with the form of the Good, our question, then, becomes how plausible is it to think that Plato’s form of the Good is incorrigible in this sense.

In asking this question, however, we immediately come up against an old chestnut concerning Plato’s commitment to propositions and propositional knowledge. If the form of the Good is not a proposition it is difficult to make sense of the question whether it is incorrigible, since the nature of incorrigibility is almost always in terms of propositions. Certainly Bailey (and Baltzly) take UFPs to be propositions. Now is obviously not the time to take up this old chestnut again. Indeed, I am not sure it is a chestnut worthy of our time for reasons I hope to pursue on another occasion. But for present purposes I want to proceed as though the chestnut has been resolved. That is, I want to proceed as though it makes sense to ask whether the form of the Good is incorrigible, either because we have found a way to account for incorrigibility (and UFPs) without essential reference to propositions or because we are comfortable thinking of the form of the Good as something like a definitional proposition. In what follows my discussion will be in terms of propositions (following both Bailey and Baltzly), but nothing (I hope) hangs on it.

---

A similar point applies to to Bailey’s intelligibility condition, which I will not be pursuing here. Indeed, I do not think that even Aristotle would accept the intelligibility condition on at least one natural reading of Bailey’s understanding of it. But since my focus in this paper is on Plato, not Aristotle, I will not pursue this thought except to say that Aristotle’s γνωστότατος (Bailey’s intelligibility) condition is ambiguous between most intelligible to us and most intelligible in nature (cf., e.g., Posterior Analytics 71b33-72a5). For a variety of reasons, I think, that Aristotle’s condition must be understood as most intelligible in nature (at least insofar as Aristotle’s UFP is analogous to Plato’s). But understood in this latter way, it does not appear to be “utterly immediate to us”; (Bailey 2006:103) emphasis added.).

Those who think the form of the Good cannot be understood propositionally will have an additional reason to reject Bailey’s and Baltzly’s accounts, insofar as they accept the conclusion of the previous section that the form of the Good is Plato’s UFP. But this, I believe, would be a mistake.
I maintain that the identity of the Plato’s UFP and the form of the Good makes suspect its incorrigibility. Plato is clear that one can be mistaken about the form of the Good. If one believes that the form of the Good is F (whether definitionally or otherwise), it does not follow that the form of the Good is F. Plato explicitly maintains that some mistakenly believe that the form of the Good is knowledge or pleasure at 505b-c, for example. Moreover, Socrates is quick to deny that, despite having beliefs about the form of the Good - indeed a belief about what the form of the Good is, he knows what the form of the Good is.²¹ That is whole point of giving the three images of the Sun, the Line, and the Cave.

Of course, this may seem uncharitable. No one would maintain that the proposition that the form of the Good is pleasure is incorrigible. Bailey’s claim must be rather that the proposition that the form of the Good is G (whatever it is) is incorrigible. The claim is that one cannot be mistaken about this latter proposition, not about the proposition that the form of the Good is pleasure. And of course, since by hypothesis the form of the Good is G (whatever it is), then one cannot believe this proposition unless it is true. But now one wonders what work the incorrigibility of the proposition that form of the Good is G is doing, beyond the proposition’s simply being true. It is difficult to see any work it could be doing other than providing a sufficient reason for stopping additional checking. But that Plato does not think that the proposition that the form of the Good is G is incorrigible in this sense is clear from the following passage.

At Republic 534b8-d1 Plato writes

Then the same applies to the good. Unless someone can distinguish in an account the form of the good from everything else (ὅς ἄν μὴ ἔχῃ διορίσασθαι τῷ λόγῳ ἀπὸ τῶν ἄλλων πάντων ἀφελών τὴν τοῦ ἀγαθοῦ ἰδέαν), can survive all refutation, as if in a battle (ἐν μάχῃ διὰ πάντων ἐλέγχων διεξέλεγχον), striving to judge things not in accordance with opinion but in accordance with being (μὴ κατὰ δόξαν ἄλλα κατ’ οὐσίαν προθυμομένους ἐλέγχειν), and can come through all this with his account still intact (ἐν πάσι τούτοις ἀπέτυχε τῷ λόγῳ διαπορεύσατα), you'll say that he doesn't know the good itself or any other good (οὔτε αὐτῷ τὸ ἀγαθὸν φήσεις εἰδέναι τὸν ὦτος ἐχοντα οὔτε ἄλλο ἀγαθὸν οὐδέν). And if he gets hold of some image of it (ἐὰν εἰδῶλου τινὸς ἐφάπτεται), you'll say that it's through opinion, not knowledge (δοξή, οὐκ ἐπιστήμη ἐφάπτεσθαι), for he is dreaming and asleep through-out his present life (τὸν νῦν βίον ὄνειροπολοῦντα καὶ ύπνώττοντα), and,

²¹See (Pappas 1974) for an argument that if a proposition, p, is incorrigible for A, then A knows that p.
before he wakes up here, he will arrive in Hades and go to sleep forever, [Grube/Reeve trans.]

Plato indicates here that having arrived at the form of the Good, which on the propositional account must be characterized as something like as having arrived at the belief that the form of the Good is G (whatever it is), one must be able to give an account of it and one must attempt to test or refute this definitional proposition (this *logos*) in every way. It is not clear what this procedure of testing or refuting is supposed to be, but whatever it is it appears that one should not be said to know the form of the Good until one’s account of it has survived these tests. Certainly, merely believing that the form of the Good is G (whatever it is) does not appear to be sufficient for concluding the inquiry, or for failing to be mistaken about it (except in the trivial way). Having gotten hold of the form of the Good one still needs to check whether one has gotten hold of the form of the Good or an image of it. Further checking is required.

If, then, the form of the Good is Plato’s UFP, Plato’s UFP does not appear to be incorrigible, at least in the sense of not requiring further checking. Having arrived at the form of the Good, further checking is required to determine whether one has in fact arrived at the form of the Good. Understood in this way, Plato does not propose an independently recognizable feature of his UFP that suffices for its truth. There is no introspective sign or intuition that suffices to bring the inquiry to a conclusion. Nor is there a recognizable incorrigibility. But if the incorrigibility of the UFP does not suffice to bring the inquiry to a conclusion what reason do we have for attributing the incorrigibility of the UFP to Plato? In asserting that we cannot be

---

22 For a beginning see (Benson forthcoming).
23 The trivial way is something like the following - if A believes that p and p is true, then p is true (or A can’t be mistaken about p). In this trivial sense, both knowledge in general and true belief are incorrigible. See *Meno* 97a-d.
24 It is true that Plato writes that one who is not able to survive these tests will be said not to have gotten hold of the form of the Good but some image of it, suggesting perhaps that one cannot get hold of the form of the Good without being able to survive these tests. There is a long story to be told here (involving among other things a return to the chestnut involving propositions), but in short Plato does not think that one can get hold of an image of F without getting something right about F (see (Benson 1992)). The point of the current passage is that one can arrive at the form of the Good (roughly, the belief that the form of the Good is G) and nevertheless, still need to check whether one has indeed arrived at the form of the Good. If this is compatible with the claim that Plato takes one’s beliefs about the form of the Good to be incorrigible, the work that incorrigibility is doing in the process of inquiry is difficult to determine.
mistaken about the UFP, Aristotle may simply be following Plato in maintaining that the UFP must be true. But not in any more elaborate way than the way in which true belief or knowledge must be true. If this is a problem for Plato’s method of inquiry, it is a problem (in the Republic at least) Plato seems willing to accept.\(^{25}\)

**Baltzly’s Account: The Negation of the Form of the Good is not Self-Refuting**

Let us turn now to Baltzly’s account of Plato’s UFP. According to Baltzly, the key feature of Plato’s (and Aristotle’s) UFP is that its denial is self-refuting in the sense that if the negation of UFP were true “neither it nor anything else could be said or thought” (Baltzly 1999:173). This too is an attempt to specify a feature of the UFP - its negation being pragmatically self-refuting\(^{26}\) - that suffices for its truth (or nearly so) and is independently recognizable.

Baltzly’s argument for his account of the nature of Plato’s UFP spans two papers. It rests primarily on the analysis of four texts (three from Plato and one from Aristotle) in which some proposition (or its negation) is shown to be pragmatically self-refuting. First, Baltzly maintains that in the first ‘consequence set’ of the Parmenides (roughly Parmenides 137-142) Plato argues that ‘One has a share of being’ by showing that ‘One does not have a share of being’ is pragmatically self-refuting.\(^{27}\) Next, Baltzly argues that at Sophist 151e, ff. Plato argues for the thesis that some kinds blend with one another and others do not by arguing that the alternatives (no kinds blend and all kinds blend) are self-refuting in the appropriate sense.\(^{28}\) Third, at Theaetetus 181c-183c Plato rejects the Heraclitean doctrine (or first principle) that everything is changing in every way by showing that it is pragmatically self-refuting.\(^{29}\) And finally, Baltzly

---

\(^{25}\)The question might be put whether Plato is committed to what has been called “the ‘pragmatist’ principle: we can only aim for what we can recognize” (see Hookway 2007:8) following Rorty. The issues surrounding Meno’s paradox suggest he is; the issues surrounding Republic 534b-d suggest he is not.

\(^{26}\)‘Pragmatic self-refutation’ is simply the name I use for the sort of self-refutation Baltzly maintains the negation of a UFP suffers from.


\(^{29}\)(Baltzly 1999:175–185), although it should be noted that Baltzly does not think the positive phase of Plato’s dialectical method is displayed here in the Theaetetus (or later in Aristotle’s Metaphysics) as it is in the Parmenides and Sophist.
turns to Aristotle’s *Metaphysics* Γ and argues that Aristotle too maintains the truth of LNC by arguing that its negation is similarly self-refuting.\(^{30}\)

Baltzly’s analysis is impressive and it is not my purpose here to take issue with his reading of these texts.\(^{31}\) Rather, my hope is simply to express concern with a key assumption of Baltzly’s account. If these passages are to inform our understanding of Plato’s method of coming to know his UFP as described in the middle books of the *Republic*,\(^{32}\) one must assume that they are best read as instances of that method. But such an assumption is problematic.

As Baltzly seems to be aware,\(^{33}\) it is difficult to square Baltzly’s interpretation of these passages with Plato’s identification of his UFP with the form of the Good. Despite insinuating that the evidence for this identification may not be as decisive as it is often thought, Baltzly concedes that for Plato “the Good is a more ultimate principle than any other” (Baltzly 1996:157). Baltzly envisions an argument to the effect that the negation of the form of the Good (presumably something like the negation of the true definitional proposition of the form of the Good) is pragmatically self-refuting, but again ultimately concedes that no such argument is to be found in the Platonic text nor, should one be found, will it “get Plato all that he wants” (Baltzly 1996:158). Such a result, of course, may not be unwelcome because, as Baltzly points out, we should be skeptical of any interpretation of Plato’s UFP which had the result that we could achieve such robust consequences (the real nature and existence of the form of the Good) from such meager resources. But in the end Baltzly concludes that either Plato never found the appropriate self-refutation argument by which he could (elenctically) establish the form of the Good as his UFP or that Plato took the method of arriving at the form of the Good to be essentially unique (and so unlike the self-refutation arguments used to establish other theses) as the Neoplatonist have suggested. I will return to this last point in a moment, but for now it

\(^{30}\) (Baltzly 1999:186–195).

\(^{31}\) See (Bailey 2006:111–119) for a critique.

\(^{32}\) That Baltzly would affirm this antecedent is indicated by the following question from (Baltzly 1996:157): “How, then, can these passages from the *Sophist* and the *Parmenides* be illustrations of the method described in Book VII in the ascent to an unhypothetical starting point?”

\(^{33}\) See the ‘Concluding Postscript’ in (Baltzly 1996:157–159), and the quote in the previous note.
suffices to note that it is difficult to square Baltzly’s account of Plato’s method for acquiring knowledge of a UFP with Plato’s method of acquiring knowledge of Plato’s UFP, the form of the Good.

Even if we put aside worries associated with the identity of Plato’s UFP with the form of the Good, other difficulties remain. Baltzly’s account of the method associated with acquiring the UFP tends to be forward looking in nature, so to speak. Baltzly looks to the *Parmenides*, *Theaetetus*, *Sophist*, and Aristotle’s *Metaphysics* to add substance to Plato’s rather sparse account of the dialectical method in the *Republic*, but all of these works are traditionally thought to post-date the composition of the *Republic*.\(^{34}\) I have no objection to looking at these works to supplement our understanding of the *Republic* passage, but in doing so we should not ignore other, perhaps more immediate and direct, evidence.\(^{35}\)

In particular, the procedure by which one acquires knowledge of the UFP, according to the *Republic*, is the method of hypothesis. In both passages in which Plato explicitly mentions the UFP, he writes about getting to it from or by means of hypotheses (510b6-7 and 511b5-7).\(^{36}\) But the method of hypothesis, as introduced and described in the *Meno* and the *Phaedo*, is composed of two procedures (as we noticed above). Having proposed a hypothesis which will help in determining an unknown answer to a question, one seeks to confirm that hypothesis \([1]\), on the one hand, by checking to see whether the consequences of that hypothesis cohere with one another, and \([2]\), on the other hand, by checking to see whether this hypothesis is a member of a set of consequences of another higher hypothesis and to see whether those consequences cohere with another. Now, there are various ways to understand these procedures and Bailey’s and

---

\(^{34}\)Nothing important hangs on this chronological claim, my primary point is that we need to take into account the *Meno* and the *Phaedo* as well, whenever any of these dialogues were written.

\(^{35}\)I grant that the *Parmenides* is at least equally direct, given Socrates’ evident reference to the method of hypothesis at *Parmenides* 135e8-136a3. The evidence of the *Metaphysics* is, perhaps, only slightly less so, given the occurrence of ανοικτοθετον, as well as other verbal echoes of the *Republic* passage. But the evidence of the *Theaetetus* and *Sophist* appears to be simply that these passages contain arguments of the same form as the arguments in the *Parmenides* and the *Metaphysics*. But again, nothing essential hangs on the directness point. My primary point is that we need to take account of the *Meno* and *Phaedo* as well, and when we do Baltzly’s account requires emendation, at least.

\(^{36}\)For a somewhat more detailed defense of this claim see (Benson forthcoming). (Bailey 2006) would seem to agree.
Baltzly’s studies are attempts to further that understanding, but on the face of it Baltzly’s account is at odds with the introduction of this method in the *Meno* and *Phaedo*.

Baltzly’s account leaves unaddressed the second confirmatory procedure - the procedure of checking for higher hypotheses.³⁷ It might be supposed that Baltzly fails to address this procedure because he is focused on the UFP, for which there is no higher hypothesis. But as we have seen above, Baltzly concedes that the form of the Good is a higher hypothesis (in his words ‘more ultimate principle’) than the UFPs Baltzly has uncovered in the *Parmenides*, *Sophist*, and *Theaetetus*. So, some account of this second confirmatory procedure ought to be included.

Furthermore, no where in any of the discussions of the method of hypothesis in the *Meno*, the *Phaedo*, or the *Republic*⁴⁶ there any suggestion that the appropriate procedure is to assume the negation of the hypothesis and check to see if it (the negation) is self-refuting. The procedure of checking to see if the consequences are coherent, which might be understood as checking see whether hypothesis is pragmatically self-refuting, is a procedure performed on the consequences of the hypothesis, not on the consequences of the negation of the hypothesis. Of course, this may be a distinction without a difference. Presumably, showing that a hypothesis is self-refuting would serve to confirm the negation of the hypothesis as a UFP, and it would hardly seem to matter whether we call the original hypothesis or its negation the hypothesis confirmed.³⁸ But this leads to another difference from the method as introduced in the *Meno* and *Phaedo*.

Once one recognizes that there is no salient difference between checking the hypothesis or its negation for self-refutation, it becomes immediately obvious that we should check both. The fact that a hypothesis (or its negation) is not appropriately self-refuting tells us nothing about

³⁷In fairness Baltzly does recognize two features of the method for ascending to UFPs. See (Baltzly 1999:173–174): “So, generally speaking, Republican dialectic has two phases: a negative phase in which the claim that ¬P is shown to be self-refuting and a positive phase in which consequences are spun out from P. These correspond to the ascent to the unhypothetical first principle and the subsequent descent through the things that depend upon it (Republic 511b7-c2).” But notice that neither of these two phases corresponds to the procedure of looking for higher hypotheses. Indeed, I have maintained elsewhere that Baltzly’s positive phase is is a constituent of the second procedure; see, for example, (Benson forthcoming). On Baltzly’s account Plato’s method of hypothesis has only one confirmation procedure (checking for pragmatic self-refutation).

³⁸This is especially true if I am right that the hypothesis and the answer to the original question are expected to be materially equivalent. See note 14 above.
whether the negation (or the hypothesis) is. That is, if Baltzly’s account of the procedure to acquiring knowledge of UFPs is correct we should expect the first confirmatory procedure to consist in checking to see whether both the hypothesis and its negation are pragmatically self-refuting. But, again, the introduction of the method of hypothesis in the *Meno, Phaedo, and Republic* indicates no such procedure. No where in the discussion of the first confirmatory procedure - the testing of coherent consequences - does Plato indicate that *both* the hypothesis and its negation should be checked for potential pragmatic self-refutation. Something like this may be suggested at *Parmenides* 135e-136c, but if it is, it looks like an emendation to the method introduced in the *Meno, Phaedo, and Republic*.

Perhaps, however, this is precisely how we should read the *Parmenides* (as well as the *Theaetetus and Sophist*) - as a more fully developed account of the method of hypothesis introduced in the *Meno, Phaedo, and Republic*, thereby accounting for the differences between Baltzly’s method and the method presented in these three earlier dialogues. As such, the method should be seen as proceeding as follows: When one seeks to come to know an answer to a question which one does not know, propose a hypothesis which will help determine that answer. Check to see whether it or its negation is pragmatically self-refuting. If one or the other is, one has hit upon a UFP and one can stop. If neither are, then check to see whether the hypothesis is a member of a set of consequences of another higher hypothesis and see whether that higher hypothesis or its negation is pragmatically self-refuting. Continue this procedure until one hits upon a UFP, i.e. a proposition whose negation is pragmatically self-refuting.

Notice that just as Bailey sought to provide a feature of the UFP that provided rational grounds for bringing to a conclusion the procedure indicated by the method of hypothesis, so Baltzly’s appeal to pragmatic self-refutation is meant to do the same thing. Hitting upon a proposition whose negation is pragmatically self-refuting provides rational grounds for accepting that proposition without attempting to derive it from higher hypotheses. But just as Bailey’s incorrigibility condition proved elusive in light of *Republic* VII 534b8-d1, so Baltzly’s self-
refutation condition will prove elusive in light of the appeal to the pragmatic nature of the self-refutation.

While Plato is interested in what Baltzly calls absolute self-refutations, which are a subset of self-contradictions, the arguments we find in the text are examples of a different sort of self-refutation - what I have been calling pragmatic self-refutation. According to this sort of self-refutation, showing that a proposition is unassertable suffices to show that it is self-refuting. But, taking self-refutation in this sense to suffice for showing that the proposition is false (and its negation true), as Baltzly points out, requires an assumption roughly to the effect that the world is optimally good and so what is true of it is assertable and/or intelligible (Baltzly 1996:157). Let us call this the assertablity/ intelligibility assumption (the AI assumption), for short. We might say, then, that the negation of a UFP is self-refuting according to Plato in the sense that the negation of a UFP is inconsistent with the AI assumption. The question becomes what is the (epistemic) status of this AI assumption. We seem to have three choices. Either this assumption is equivalent to (or a constituent of) the form of the Good, or it is distinct from the form of the Good, but is a UFP, or it is not a UFP.

Let us suppose Plato took this AI assumption to be equivalent to (or a constituent of) the form of the Good. If so, then the method of arriving at the form of the Good will be essentially different from the method of arriving at other Platonic UFPs on Baltzly’s own account. It will do no good to show that the negation of the form of the Good is inconsistent with the AI assumption, since the negation of the form of the Good just is the denial of the AI assumption. Either the negation of the form of the Good must be thought to be absolutely self-refuting or the method of arriving at the form of the Good must be importantly different than showing its negation is self-refuting in either of these senses. But there is no reason to think that the AI assumption is absolutely self-refuting (or at least neither Baltzly nor Bailey have given us any

39See (Baltzly 1996:162–163 n. 19), although Baltzly provides no evidence for Plato’s interest.
40Exactly how this is supposed to work, I will not pursue, since for the reasons given below I think this supposition like the next is a non-starter.
reason to think it is). Moreover, it is exceedingly odd that the only proposition that we have arguably direct evidence that Plato takes to be a UFP is arrived at by means of a different method than the one Baltzly claims to be the method by which Plato thinks we arrive at a UFP. So, supposing that the AI assumption is the form of the Good looks like a non-starter.

Let us suppose, then, that the AI assumption is not the form of the Good, but is, nevertheless, a UFP. This doesn’t help. The AI assumption is not reached by showing its negation is self-refuting in the way that the negations of the other UFPs are self-refuting, since the negation of this assumption just is the denial of the AI assumption. So either this assumption must be reached by discovering that its negation is absolutely self-refuting, which we noted above we have no reason to think that it is, or it must be reached by a method different than Baltzly’s method of reaching UFPs. What is worse, the method by which the form of the Good (as another UFP) remains undetermined. Again, the supposition that the AI assumption is a UFP distinct from the form of the Good is a non-starter.

This leaves us supposing that the AI assumption is not a UFP. But if it is not, then why on Platonic grounds should we think it is true? Indeed, it would seem that the AI assumption is in need of confirmation like any other hypothesis proposed in the course of the employment of the method of hypothesis. We should expect that one will need to confirm it by checking whether its consequences cohere with one another (in some sense other than absolute self-refutation or pragmatic refutation), and one will need to look for a higher hypothesis whose consequence the AI assumption is and test that higher hypothesis for coherent consequences. But insofar as the truth of the AI assumption is open to question, so the truth of the hypothesis whose negation is inconsistent with the AI assumption is open to question. That is, coming to believe a proposition whose negation is inconsistent with the AI assumption (i.e. which is pragmatically

---

41 Of course, Plato might have mistakenly thought it was. But if so, we should like some evidence for such an uncharitable attribution.

42 See the comment about the Neoplatonists above. Note that the other UFPs Baltzly attributes to Plato are thought to be UFPs only because they are arrived at by means of the method that Baltzly argues is the method by which Plato thinks UFPs are reached.
self-refuting) does not guarantee the truth of that proposition, any more than coming to believe a proposition whose negation is inconsistent with the proposition that virtue is good. Consequently, method of testing whether a proposition’s negation is pragmatically self-refuting is not essentially different than testing whether a proposition’s negation is inconsistent with other things one happens to believe.43

Baltzly, then, has failed to identify a unique feature of the method of obtaining Plato’s UFP. The feature of having a negation that is pragmatically self-refuting does not apply to Plato’s form of the Good, the only the thing for which we have any independent evidence as being a Platonic UFP. Moreover, such a feature, insofar as it does apply to Platonic propositions is not substantially different than the features that apply to any hypothesis that has not yet been confirmed all the way up to Plato’s UFP, i.e, the form of the Good. Nevertheless, Baltzly has identified a feature of some propositions/ hypotheses that evidently attracted Plato’s (and Aristotle’s) attention, viz. the proposition’s/ hypothesis’ negation is inconsistent with the assertability/ intelligibility assumption (AI). But short of evidence that Plato takes this feature to have some special epistemic salience, Baltzly (like Bailey) has failed to identify an independently recognizable feature of Platonic UFPs (or even their near relatives) that is sufficient for recognizing their truth.

Conclusion

Both Bailey and Baltzly have offered sophisticated accounts of Plato’s UFP based in part on Aristotle’s account of his UFP in *Metaphysics* Γ. There is much we and should be learned from their discussions. But in the end, whatever Aristotle may believe about his UFP (and I genuinely mean to leave this as an open question), either the evidence most directly relevant to Plato’s discussion of his UFP fails to reveal the feature of his UFP which allows one to recognize it as the UFP that it is, or the evidence we have (*Republic* 534b8-d1) indicates that according to Plato there is no feature of his UFP that allows us to recognize it as the UFP that it is. If this is a

43If I am right here, then Baltzly has failed to show that what I am calling pragmatic self-refutation is importantly different than dialectical refutation; see (Baltzly 1996:162 n.18).
problem for Plato’s method of inquiry, (as I said earlier) it is a problem that Plato is unable to resolve or fails to reveal to us how he does resolve it.