

## University of Groningen

### Plato on pleasure and illusion

van Zoonen, Derek

DOI:  
[10.33612/diss.250286363](https://doi.org/10.33612/diss.250286363)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2022

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*  
van Zoonen, D. (2022). *Plato on pleasure and illusion*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.250286363>

#### **Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### **Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

### 3. ‘The Greatest Impostor’: Hedonic Cognitivism, Hedonic Fallibilism, and Deceptive Pleasure in Plato’s *Philebus* (36c3–41a4)

In the middle of the *Philebus* (36c3–51a1), the dialogue in which Plato presents his final and most developed thoughts on the subject of the nature and value of pleasure and its place in a good life, Plato’s Socrates ‘stirs up a weighty controversy (λόγον οὐ πάνυ μικρόν, 36d3–4)’<sup>246</sup> by suggesting that pleasures can be ἀληθής (true, real) and ψευδής (false, unreal, deceptive). Such falsity, he continues, comes in four types of which the first type will concern us here—the pleasures that are false in the same, literal, sense in which our beliefs can be false because they fail to represent the world correctly.<sup>247</sup>

Already in antiquity, Plato’s readers were perplexed, if not annoyed by this talk of false pleasure. Aristotle’s successor in the Peripatetic school, Theophrastus, argued that ‘if there is such a thing as false pleasure, there will be pleasure that is not pleasure’ and concluded that all pleasures are true,<sup>248</sup> Gallop complained that Plato’s doctrine of false pleasure is so muddled that it ‘annually defeats my

---

<sup>246</sup> Throughout this chapter, references are to the *Philebus* (unless otherwise noted) and I more or less follow Frede’s (1993) translation of the Greek provided by Burnet (1903) with small modifications here and there.

<sup>247</sup> From here on, I will simply refer to this first type of false pleasures as false pleasure. The other false pleasures under discussion are, in order of appearance, pleasures that are false because they seem larger or more intense than they really are (41b4–42c4); pleasures that are false because they are in fact neutral states pretending to be pleasures (42c5–44d6); and—finally—pleasures that are false because they are not pure but intimately mixed up with pain (44d7–50e4). For a neat and accessible overview, see e.g. Frede (1992); for an attempt to bring some order in these different types of falsity, see Fletcher (2018-b) and Strohl (unpublished manuscript).

<sup>248</sup> Theophrastus fragment 85 (Wimmer) cited by Gallop (1960: 331).

powers of exposition,<sup>249</sup> and Gosling sparked an enormous debate when he accused Plato of the confusing different senses of truth and falsity or ‘rank equivocation.’<sup>250</sup>

To be sure, Plato’s defence of the possibility of false pleasures or hedonic mistakes is highly puzzling. It is especially unclear, firstly, how a private and subjective mental state like pleasure—arguably nothing more than a bit of feeling—could be false in precisely the same way in which a belief can be false. Untutored intuition, here represented by Protarchus who takes up Philebus’s unadulterated hedonism, resists this proposal for the simple reason that, unlike representational states (such as beliefs or perceptions), an affective state does not constitute an experience ‘in which the world reveals itself to us’ as McDowell puts it.<sup>251</sup> Instead, the thought goes, pleasure belongs to the phenomenal realm to which we have infallible or at least privileged access.

It is also unclear, secondly, what role Plato’s defence of hedonic fallibilism plays in the overall economy of the *Philebus*. What is the connection, one might ask, between the possibility of hedonic mistakes or false pleasures and Socrates’s and Protarchus’s attempt to discover some ‘state and condition of the soul . . . that can give all human beings a happy life (ἔξιν ψυχῆς καὶ διάθεσιν ... τὴν δυναμένην ἀνθρώποις πᾶσι τὸν βίον εὐδαίμονα παρέχειν, 11d4–6)?’

My main aim in this chapter is to untangle—or at least elucidate—these puzzles, especially the first. In brief, I am going to argue that Plato’s defence of hedonic fallibilism is predicated upon a

---

<sup>249</sup> Gallop (1960: 338).

<sup>250</sup> See Gosling (1959) and Gosling (1975). For a rejoinder, see Fletcher (2018-b) and Szaif (2021: 136).

<sup>251</sup> McDowell (1998: 330).

position I call hedonic cognitivism.<sup>252</sup> On this theory of pleasure, hedonic states are more than blind surges of affect, raw sensations, or mere feelings. Put bluntly, Plato argues that pleasure is a way of taking the world to be a certain way. If pleasure is in the business of representing reality, though, it straightforwardly follows that a pleasure can be true or false: like other representational attitudes, a pleasure is true if its content represents some way the world really is and it is false if its content *misrepresents* some way the world really is.

With this discussion of the first puzzle in place, I turn to the second puzzle and explain how the possibility of false pleasure bears on the dialogue's structuring question—specifically focusing on Socrates's mysterious claim that false pleasures are 'ridiculous imitations (μεμμημέναι τὰς ἀληθεῖς ἐπὶ τὰ γελοιώτερα, 40c5)' of true pleasure. False pleasure, I argue, is deceptive pleasure: the subject of the false pleasure is in the grip of a misrepresentation of the world and thus out of contact with reality even though this cannot be appreciated from an inside perspective because, like a false belief, a misleading pleasure necessarily represents its content as true.

This chapter is organized in the following way. Having briefly introduced the *Philebus* (section 1), I offer a close examination of Plato's argument for false pleasure which I dub the Fallibilism Argument and which runs from 36c3–41a4 (section 2). This argument, I suggest, is best read as a dialectical back and forth between Socrates and Protarchus in which they defend their diametrically opposed takes on pleasure. I contend that, throughout the argument, Socrates finetunes his hedonic fallibilism, as well as the underlying theory of pleasure, in response to Protarchus' infallibilist objections.

---

<sup>252</sup> Hedonic cognitivism should be distinguished from a weaker thesis concerning the connection between pleasure and cognition: the claim—readily accepted by Protarchus—that pleasure is often accompanied by, or associated with, a (potentially false) belief.

In the first move—the First Hedonic Cognitivism Argument (37a1–e7)—Socrates more or less stipulates that, like beliefs (δόξαι), pleasures have representational content in virtue of which they can be evaluated in terms of truth, accuracy, or correctness. When Protarchus resists this argument by claiming that pleasure is immune to error, Socrates surprisingly enough starts to defend a weaker theory than his original hedonic cognitivism. On this weaker theory, our pleasures have (derivative) semantic value because they are *accompanied by* or *associated with* true or false beliefs.

When Protarchus resists this proposal by pointing out that, in this case, it is merely the *belief* that is false and not the pleasure, Socrates returns to his stronger line of argumentation and claims that there is a tighter connection between the cognitive element and the pleasure. This line of argumentation—the Second Hedonic Cognitivism Argument (38b6–41a4)—is best understood, I argue, as a more careful attempt to explicate what the earlier defence of hedonic cognitivism had left implicit.

More specifically, the argument explains where pleasures derive their content from. Its basic suggestion is that pleasures can be false because they are *constituted by* representational or cognitive elements. Drawing on an elaborate metaphor of the human psyche—and mainly focusing on future-oriented, anticipatory pleasures in whose case it is clear that representational elements play a crucial role—Socrates convinces Protarchus that pleasures are more than mere bits of experience: they centrally involve representations or cognitions (φαντάσματα based on δόξαι) of some state of the world. Since any representation can misfire, it follows that pleasures can be true (if they represent the world correctly) or false (if they misrepresent the world).

In section 3, I turn to a thorny issue and seek to say more about the precise nature of this falsity—situations, that is, where pleasure

misrepresents—building on extant scholarship. Even if we grant Socrates that a pleasure can be false thanks to its false content, what is it that gets misrepresented by such false pleasure? Much of the existing literature on this thorny issue can be divided into descriptive and evaluative readings. According to the descriptive or factual interpretation, a pleasure taken in some descriptive state of affairs *p* is false if *p* is not the case; according to the evaluative or ethical reading, by contrast, a pleasure taken in *p* is false because one considers *p* to be good (or otherwise positively evaluatively charged), although *p* is *not* good (or otherwise negatively evaluatively charged).

I resist this stark dichotomy for a couple of reasons. For one, it strikes me that descriptive and the evaluative readings share an important common ground that is often overlooked or obscured in the literature.<sup>253</sup> In both cases, I maintain, false pleasure is simply a matter of *misrepresentation*: whenever I experience false pleasure, no matter whether we construe this falsity descriptively or evaluatively, there is mismatch between my representation of the world and how that world actually is. Being in the grip of false pleasure is always a matter, then, of being out of contact with reality.

For another, I believe there is good reason to believe that, like other emotions or passions (*πάθη*), pleasure involves both a factual or descriptive and an evaluative or ethical cognitive element. What the interpretive dichotomy misses is that any pleasure necessarily involves a factual belief that picks out an object or state of affairs in the world and an evaluative belief—an appraisal or construal, as psychologists call it—that ascribes positive value to it. Without these two beliefs, a pleasure cannot get off the ground.

With this account in place, I move to section 4 where I draw some conclusions and argue that this more complex, more

---

<sup>253</sup> E.g. by Evans (2008-a) and by most evaluative readings which have been designed in conscious opposition to the traditional factual reading.

sophisticated picture of pleasure fits neatly into the overall aim of the *Philebus* and is intimately connected with the attempt to discover what a good life looks like—a connection that has received surprising little attention in the literature. Plato’s hedonic cognitivism is not just a perfect manifestation of one of the central findings of the *Philebus*—the idea that cognizance plays a necessary role in any aspect of a life well lived, including pleasure—it also makes trouble for the unmitigated hedonism of Protarchus and Philebus in a variety of ways.

### 1. The Philebus, Pleasure, and the Authority of the Soul

In the *Philebus*, Socrates and the hedonist Protarchus set themselves the task of ‘trying to prove some possession or state of the soul to be the one that renders life happy for human beings (ἔξιν ψυχῆς καὶ διάθεσιν ... τὴν δυναμένην ἀνθρώποις πᾶσι τὸν βίον εὐδαίμονα παρέχειν, 11d4–6).’ While Protarchus takes pleasure (ἡδονή), delight (τέρψις), or enjoyment (τὸ χαίρειν) to be the state of the soul they are looking for, Socrates opts for knowledge (ἐπιστήμη), reason (νοῦς), or intelligence (φρόνησις).

Upon closer scrutiny, however, Socrates and Protarchus discover that they are both wrong: neither the life of pleasure nor the life of knowledge but rather the mixed life in which the two are combined turns out to be the good life (20b1–23b4). With this reoriented understanding of what makes for a good life, the guiding question of the dialogue as a whole shifts significantly too: no longer are Socrates and Protarchus interested in either pleasure or knowledge, what drives them from here on is the attempt to find out which of the two is responsible for the fact that the mixed life counts as the good life.

It is obvious, though, that before one can compare the influence

of two ingredients on the overall quality of a mixture, one should first know what these ingredients precisely are and what conditions or circumstances bring them about. To find this out, Socrates thinks they need ‘a satisfactory examination’ (ικανῶς βασανίσαι) of especially pleasure, and to develop such a satisfactory examination, in turn, they need to ‘find out in what kind of thing each of them resides and what kind of condition makes them come to be when they do (ἐν ᾧ τέ ἐστιν ἐκάτερον αὐτοῖν καὶ διὰ τί πάθος γίγνεσθον ὅποταν γίγνησθον, 31b2–5).’

A promising way to understand Socrates’s strategy here is to assume that he aims to establish ‘the authority of the soul’ (to borrow Harte’s felicitous phrase) in all affective matters.<sup>254</sup> The basic point, in other words, is that our affective experiences cannot be understood in way independent from cognizance and other mental activities or operations.<sup>255</sup> Even raw physical pleasures—the pleasure of getting a massage, for instance, or the pleasure of drinking cold water on a hot summer day—only count as pleasures because they enter conscious awareness and get picked up on by our psyche. In an important way, then, there are strictly speaking no such things as purely bodily pleasures: all pleasures heavily depend on the mind.

What is more, a second type of pleasure is solely psychological or mental. These anticipatory pleasures, as they are often called, turn out to be ‘a different kind of pleasure (ἕτερον εἶδος)’ and they are intimately connected with ‘the anticipation that the soul experiences by itself without the body (τὸ χωρὶς τοῦ σώματος αὐτῆς τῆς ψυχῆς διὰ προσδοκίας γιγνόμενον, 32c3–5).’ Think for instance of the pleasure I feel *right now* because I am fantasizing about next week when, if all goes well, I will be enjoying a glass of slightly cooled pinot noir while sitting in the warm French sun.

---

<sup>254</sup> Harte (2014).

<sup>255</sup> See e.g. Harte (2014) and J. E. Butler (2007).



To finetune his analysis of this type of pleasure, which plays an essential role in his subsequent defence of the possibility of false pleasure, Socrates offers a more careful examination of the psychology undergirding our affective experiences in which he claims that especially memory and desire play a pivotal role in the genesis of such anticipatory pleasures. When we are in a state of lack, such as being hungry, our mind is aware of this lack and draws on memories of past events to have us realise that there is a state opposed to being hungry. This realization, in turn, leads to a desire for this opposite state. As Socrates has it, 'every living creature always strives towards the opposite of its own experience (τοις ἐκείνου παθήμασιν ἐναντίαν ἀεὶ παντὸς ζῴου μηνύει τὴν ἐπιχείρησιν, 35c9–10).'

After this excursus into the psychology underlying our affective experiences, the pleasures of anticipation enter the scene again. An 'empty' person (κενούμενος), as Socrates formulates it at 36b4–6, sometimes 'hopes or expects' that he will be 'filled (ἐν ἐλπίδι τοῦ πληρωθήσεσθαι)' in the near future. When this happens, this person takes pleasure in the anticipation of being filled again: he or she enjoys the idea of future enjoyment or takes pleasure in the idea of prospective pleasure.

## 2. The Hedonic Fallibilism Argument

At this point, Socrates suddenly and rather unexpectedly introduces the notion of *falsity*: he explicitly says that he wants to make use of their 'examination of those affections for a particular purpose (ταύτη δὴ τῇ σκέψει τούτων τῶν παθημάτων τόδε χρῆσώμεθα, 36c)'—namely to see whether the concept of falsity can be applied to

pleasures.<sup>256</sup> As the passage continues:

[T3.1] [Socrates] Shall we say that those pleasures and pains (ταύτας τὰς λύπας τε καὶ ἡδονὰς) are true (ἀληθεῖς) or false (ψευδεῖς), or that some are true and others not so (τὰς μὲν τινὰς ἀληθεῖς, τὰς δ' οὐ)?

[Protarchus] But, Socrates, how can there be false pleasures or pains (πῶς ... ἂν εἶεν ψευδεῖς ἡδοναὶ ἢ λύπαι)?

[S.] But, Protarchus, how can there be true and false fears (φόβοι), or true and false expectations (προσδοκίαι), or true and false beliefs (δόξαι)?

[P.] Beliefs (δόξας) I would grant you, but not the rest.

[S.] What? I am afraid we are starting a very considerable discussion (λόγον οὐ πάνυ μικρόν).

[P.] You are right. (36c6–d5)

There are, I think, a few important things in this passage that are worth emphasizing. As Gosling and Taylor have put it, 'a major source of obscurity about this topic is, precisely, what the topic is.'<sup>257</sup> It is not clear, for instance, what the scope of Socrates's claim is, how we should unpack the notion of falsity Socrates is working with here, or what Protarchus is suggesting when he pushes back against the bold proposal that pleasures and pains can be false.

To answer these and other questions, I think it pays off to focus on Protarchus's ultimate conversion to the thesis that pleasures can be false. What happens in the handful of Stephanus pages which make up the Fallibilism Argument that brings Protarchus to concede at 40b5 and 40c7 that false pleasures do indeed exist, although at 36c8–9, 36e9–10, and 37e12–38a2 he had vehemently disagreed with Socrates's

---

<sup>256</sup> Following Whiting (2014: 24), we can say—as a rough first pass—that a pleasure is false just in case it misleads us about some way the world really is.

<sup>257</sup> Gosling and Taylor (1982: 429).

contention that there are false pleasures?<sup>258</sup>

If we look at the exchange between Socrates and Protarchus from 36c3 up to 41a4, I think we can discern three different stages in all three of which there is a dialectical back and forth between Protarchus's infallibilism and Socrates's fallibilism. [1a] When Socrates broaches the notion of true and false pleasures, Protarchus presents and defends a first iteration of his hedonic infallibilism or what I will call his Immunity To Error Thesis. [1b] In response to Protarchus's infallibilism, Socrates offers an argument from analogy which I will call his First Hedonic Cognitivism Argument. This argument more or less stipulates that, like δόξα, pleasure has content. Seeing as a δόξα can be true or false in virtue of its content, it follows that a pleasure can be true or false too in virtue of its content. It is not the case, in other words, that 'pleasures are simply what they are (ἡδονὴ δὲ καὶ λύπη μόνον ἄπερ ἐστὶ, 37c5)': they can be evaluated in terms of truth, correctness, rightness, properness, or accurateness.

[2a] Protarchus pushes back again and offers a slightly different formulation of his infallibilism which claims that 'pleasure cannot be mistaken (οὐχ οἷόν τε, εἴπερ ἀμαρτήσεται γέ ἡδονή, 37e8–9).' [2b] Mysteriously enough, Socrates suddenly weakens his proposal to get Protarchus on his side: instead of arguing that pleasures have truth-evaluable content, as he did before, he now claims that our pleasures are often accompanied by beliefs with truth-evaluable (or false) content.

[3a] When confronted with yet another iteration of Protarchus's Immunity To Error Thesis ('in [Socrates's alleged false pleasure case] the belief is false, no one would dream of calling the pleasure false (τὴν μὲν δόξαν γέ ἐν τῷ τοιούτῳ καὶ τότε λέγομεν

---

<sup>258</sup> Cf. Guthrie (1978: 219): '[A]t 36d and 38a Protarchus denies the existence of false pleasures. At 40b–c he admits it. So Socrates must have said something in between to make him change his mind.'

ψευδῆ, τὴν δ' ἡδονὴν αὐτὴν οὐδεὶς ἄν ποτε προσείποι ψευδῆ, 37e12–38a2), [3b] Socrates's Second Hedonic Cognitivism Argument (38b6–41a4) returns to the stronger proposal driving the First Hedonic Cognitivism Argument (37a1–e7) and strengthens it. To do so, Socrates fleshes out his earlier suggestion that pleasures have content (instead of merely stipulating it) and focuses on future-oriented pleasure (which makes it easier to see that we enjoy our representations of the world rather than the world itself). [3c] Protarchus finally gives up his immunity to error thesis and agrees with Socrates that 'there are false pleasures in human souls (εἰσὶν ... ψευδεῖς ἐν ταῖς τῶν ἀνθρώπων ψυχαῖς ἡδοναί, 40c4–6).' These stages and their central questions are best discussed in turn.

## 2.1 Protarchus's Immunity To Error Thesis

Let us start with Protarchus' first formulation of his Immunity To Error Thesis, to be found in [T3.1]. Although Protarchus initially denies that 'those pleasures and pains' (ταύτας τὰς λύπας τε καὶ ἡδονὰς)—namely the anticipatory pleasures and pains they have just been talking about—can be false, a little later his denial seems to become a more sweeping denial of false pleasures and pains *as such* or *in general*.<sup>259</sup>

There are two ways of unpacking his denial of false pleasures. On a first construal, Protarchus's claim is that there are no true or false pleasures. Put in another way, he is denying, on this reading, that affective states are *truth-apt*. On an alternative construal, he is claiming that all pleasures are true or that there are no false pleasures. At least

---

<sup>259</sup> Cf. the exchange between him and Socrates at 36e2–4 ('What is your view? Are there not false pleasures as well as true ones? But how could there be?') and 40d7–10.

since Kenny,<sup>260</sup> it has been recognized that Protarchus's rejection of false pleasure does not imply the rejection that pleasure can be truth-apt: from the start it is clear that Protarchus has no qualms about pleasures being called *true*, he only resists their being called *false*.

Initial, albeit non-conclusive evidence can be found in [T3.1] where Protarchus only wonders how there could be false pleasures and pains.<sup>261</sup> Additional evidence can be distilled from the following exchange:

[T3.2] [Socrates] [W]hat we have to question is how it is that belief (δόξα) is usually either true or false (ψευδής τε καὶ ἀληθής), while pleasure admits only truth (τὸ τῆς ἡδονῆς μόνον ἀληθές), even though in both cases one is equally really believing and taking pleasure (δοξάζειν ὄντως καὶ χαίρειν ἀμφοτέρω ὁμοίως εἴληχεν).

[Protarchus] We have to question that. (37b5–9)

Kenny seems right, then, to conclude that '[n]owhere is [Protarchus] made to deny that pleasures can be true; nowhere does he suggest that pleasure has no true-false poles at all.'<sup>262</sup>

Even after some pressure from Socrates—who introduces extreme circumstances like dreams, madness, and delusions to convince Protarchus' of the possibility of false pleasures—Protarchus holds his ground, asks the rhetorical question 'how could pleasures be false (πῶς γὰρ ἄν)?', and retorts that 'we all hold the view that all these things are so (πάνθ' οὕτω ταῦτα ἔχειν πάντες ὑπειλήφαμεν), Socrates (36e9–10).' Here, then, is an attempt to capture the first

---

<sup>260</sup> Kenny (1960: 47); cf. Wolfsdorf (unpublished manuscript: 3).

<sup>261</sup> Cf. 38a2 where Protarchus merely denies that pleasures can be false.

<sup>262</sup> Kenny (1960: 47). He continues: 'To us it may seem clear that if it does not make sense to call something false, it does not make sense to call it true either [...], but perhaps such things were not always so clear.'

## iteration of Protarchus's Immunity To Error Thesis:

All pleasures are true; there are no false pleasures.

Although the literature contains more than one way of understanding this thesis and Protarchus's reasons for holding it,<sup>263</sup> one promising way to make sense of it that I will broadly follow here is to say that Protarchus is committed to something like a Cartesian transparency thesis or the idea that our passions or emotions 'are so close and internal to our soul that it is impossible that it should feel them without their truly being as it feels them.'<sup>264</sup> All pleasures are true—in the sense that we cannot be in doubt, ignorance, or error about them—simply because we have the privilege of first-person, immediate access to mental states like pleasure and pain.<sup>265</sup>

---

<sup>263</sup> For good discussion, see Wolfsdorf (unpublished manuscript). Some other options include: Protarchus believes that a pleasure is true insofar as that in which it is taken is truly pleasant and a thing being truly pleasant is a function of one's finding it so (Harte); Protarchus's commitment to the Immunity To Error Thesis is more or less obscure (Gosling and Taylor); Protarchus thinks (just like Theophrastus) that there are no non-genuine pleasures (Kenny); Protarchus resists the view that pleasure can be false because if he accepts this view, he must concede to Socrates that some pleasures are not good (Frede); Protarchus is a 'Protagorean hedonist' who believes—as Mooradian (1996: 94) puts it—that 'pleasure cannot be false because it is a kind of αἴσθησις and αἴσθησις is always correct in relation to its objects' (Mooradian and Warren).

<sup>264</sup> Descartes, *Les passions de l'âme*, article 26. In his commentary on the *Philebus*, Delcomminette treats self-ascription of pleasure as a kind of Cartesian *cogito*: 'Ne touchons-nous pas ici à l'évidence ultime, celle du *cogito* lui-même?'

<sup>265</sup> Cf. Frede (1997: 245): 'Auch mit seinem nächsten Vorschlag, man könne doch im Traum oder im Wahnsinn fälschlich meinen, Lust oder Unlust zu empfinden, ist Sokrates nicht erfolgreicher. Protarchos bestreitet, daß es sich dabei um 'falsche' Lust oder Unlust handelt. Er scheint davon auszugehen, daß man sich über seine Gefühle gar nicht im Irrtum befinden kann.'

When talking about subjective experiences, the thought goes, we have as it were reached rock bottom with regard to justification — there is no point in denying honest, immediate self-reports about the nature of our affective (and other subjective) experiences.<sup>266</sup> This, in its turn, might have to do with the fact that there is no appearance reality when it comes to pleasure: the essence of pleasure consists in being felt. As Socrates reformulates Protarchus’s claim, pleasures (and pains) are ‘simply what they are (ἡδονὴ δὲ καὶ λύπη μόνον ἄπερ ἔστί, 37c5).’

This construal of Protarchus’s commitment to the Immunity To Error Thesis ties in neatly with the (admittedly speculative) assumption that Plato might have used Philebus and Protarchus—the first of whom is almost certainly a literary fiction—to give voice to hedonistic ideas that were fashionable in Plato’s lifetime in the Cyrenaic school.<sup>267</sup> The Cyrenaics were famous for defending a radically subjectivist epistemology, according to which sensible qualities and sensible objects are nothing but motions or modifications of our consciousness (πάθη). They also believed that these modifications of consciousness are either painful, pleasurable, or indifferent, that they do not allow for any further classification, and

---

<sup>266</sup> Much (recent) empirical and philosophical work, which claims that we can be in serious error even about our subjective experiences, obviously makes trouble for far-reaching Cartesian infallibilism. See e.g. Armstrong (1963), Bayne and Spener (2010), Churchland (1988), Haybron (2007), Haybron (2008), Jäger (2009), W. James (1884), Schwitzgebel (2008), and Wilson (2002).

<sup>267</sup> As, most recently, Zilioli (2014: 83–5) has argued. Some remarkable similarities are, for instance, the following ideas: that we are infallible and incorrigible about our mental states, that pleasure is the *summum bonum* for all living beings, that is, that the pleasant is good and that the good is pleasant, that pleasure does not differ in kind but only in degree so that we cannot apply any other predicate to it except that it is pleasurable, and the conviction that pleasure is good even if it proceeds from bad or immoral conduct. For a discussion of Cyrenaic philosophy, see e.g. Tsouna (2005), Lampe (2014), and Zilioli (2014).

that we are infallible and incorrigible when it comes to (our reports about) these πάθη. (They are always true, as Protarchus would put it.) Whenever I candidly report that I am undergoing some affection, the Cyrenaics held, I do so ‘infallibly and truly and firmly and incorrigibly (ἀδιαψεύστως καὶ ἀληθῶς καὶ βεβαίως <καὶ> ἀνεξελέγκτως)’ (as Sextus Empiricus, *Against the Mathematicians* 7.191 reports).<sup>268</sup>

## 2.2. The First Hedonic Cognitivism Argument (37a1–e7)

The First Hedonic Cognitivism Argument (37a1–e7), which pivots on an analogy between pleasure and belief (δόξα), moves in three simple steps and goes something like this:

1. A belief is false if its content misrepresents a state of the world.
2. Like a belief, a pleasure has content.
3. So a pleasure is false if its content misrepresents a state of the world.

To make his case, Socrates starts by breaking down pleasure (ἡδονή) and belief (δόξα) in two parts or aspects to show that both are structurally similar. Belief, on the one hand, can be divided into the act of believing (δοξάζειν, 37a2) and what is believed (τὸ δοξαζόμενον, 37a7). Pleasure, on the other hand, can similarly be divided into the act of taking pleasure (ἡδεσθαι, 37a5) and—to

---

<sup>268</sup> Cf. *Theaetetus* 152b2–5 for Socrates’s ascription of a similar idea to Protagoras *cum suis*.



formulate it somewhat awkwardly—‘what that which takes pleasure in takes pleasure in’, ‘that in virtue of which what takes pleasure takes pleasure’, or ‘that in or by which what takes pleasure takes pleasure’ (τό ᾧ τὸ ἡδόμενον ἦδεται, 37a9).

What do these relatively vague expressions τὸ δοξαζόμενον and τό ᾧ τὸ ἡδόμενον ἦδεται refer to? It is often simply assumed—for instance in Frede’s translation—that these expressions refer to the *object* of our beliefs and pleasure, but this interpretation is unsatisfactory for a couple of reasons. Most importantly, as Delcominette has pointed out, this construal makes it very difficult, if not impossible to understand how hedonic error is possible. ‘[F]or if τὸ δοξαζόμενον corresponds to the object of belief, what could it mean that this object is *mistaken*? One should rather say that the subject is mistaken *about* the object.’<sup>269</sup> The expression τὸ δοξαζόμενον does not refer to the *object* of the belief, then, but to its *content*. This content can be said to be mistaken, or false, if it does not correspond adequately to what the δόξα is about—and the same would hold for pleasure.

Like belief,<sup>270</sup> Socrates’s thought seems to go, pleasure is a unitive yet double-edged complex which consists of two inseparable

---

<sup>269</sup> Delcominette (2003: 217). Here are two more worries for this interpretation. Firstly, note that when Socrates wants to draw a distinction between subject and object rather than a distinction between the act of believing or enjoying and its content, he uses the expressions ‘what is believing’ and ‘what is enjoying’ (37a–b) to refer to the *subject* of these acts. These expressions are not used here which seemingly suggests that we must be dealing with an act-content rather than a subject-object opposition. Secondly, Socrates goes on to use these distinctions to analyze future-oriented or anticipatory pleasures in which case there might be no objects to be referred to—even though there *is* content.

<sup>270</sup> Both in English and in Ancient Greek, the word belief or δόξα is ambiguous: it refers both to the *attitude* we have when taking something to be true as well as to *what* we take to be true (the content of the attitude).

aspects or faces: for any pleasure, there is the brute fact *that* we are taking pleasure in something (the act) and there is the *content* of this attitude (*what* we are taking pleasure in) which specifies it as this or that pleasure. Importantly, it is the content that is responsible for the truth or falsity of the belief or pleasure as a whole.<sup>271</sup>

This reading is indeed borne out by the text. We see that Socrates agrees with Protarchus that that which believes (τὸ δοξάζον) cannot be deprived of (ἀπόλλυσιν) the fact that it really believes (τό δοξάζειν ὄντως), whether it believes correctly (ὀρθῶς) or not. Even though the content of my belief can be mistaken, in other words, it makes no sense to say that the act of believing—the attitude I have while taking something to be true—gets things wrong (37a11–b1). Similarly, Socrates continues, that which takes pleasure (τὸ ἡδόμενον), whether it takes pleasure rightly (ὀρθῶς) or not, can never be deprived of (ἀπόλλυσιν) the fact that it really takes pleasure (τό ὄντως ἡδέσθαι) (37b2–4).

This suggests that the possibility of error or mistake must be sought in the *content* of pleasure and belief rather than in the acts of believing or being pleased themselves. Our mental acts cannot be false—we are, so to speak, still believing and taking pleasure even if the content of our doxastic or hedonic attitude is false—so if we want to ascribe falsity to our pleasures and beliefs as a whole, we must ascribe falsity to their content in virtue of the fact that it does not adequately represent reality. In this respect the content of our beliefs and pleasures can either display rightness (ὀρθότης) or some opposite of rightness (τοῦναντίου ὀρθότητι), reach their aim or miss it (ἀμαρτάνειν).<sup>272</sup> More precisely, if the content of the belief gets things

---

<sup>271</sup> Delcomminette (2003: 218).

<sup>272</sup> In the *Cratylus* (420b–c), Socrates offers an etymology of δόξα that traces it back to the word τόξον: this suggests that belief is like a bow aimed at an object which means, in turn, that, like archers, our beliefs can hit their target or miss the mark. The same

wrong, we say that the belief—as a whole—is incorrect or false. Analogously, if pleasure is mistaken (ἀμαρτάνουσαν) in what it is pleased about, we should not call it right (ὀρθήν) or proper (χρηστήν) or apply other names of praise (τι τῶν καλῶν ὀνομάτων) (37d6–e7). Instead, such a pleasure is *false*.<sup>273</sup>

### 2.3 Socrates Defends a Weaker Position

Protarchus demurs, however, and repeats his Immunity To Error Thesis in a somewhat different formulation: ‘That would be impossible, if indeed (γε) pleasure (ἡδονή) should be mistaken (ἀμαρτήσεταιί, 37e8–9).’ Again, Socrates tries to undermine Protarchus’s position, but still without success:

[T3.3] [Socrates] As to pleasure, it certainly often seems to arise (γίγνεσθαι) in us not with a right (οὐ μετὰ δόξης ὀρθῆς) but with a false belief (μετὰ ψεύδους).

[Protarchus] Of course. But what we call false (ψευδῆ) in this case (ἐν τῷ τοιούτῳ) at that point (τότε) is the belief (τὴν δόξαν), Socrates; nobody would dream of calling the pleasure itself false (τὴν δ’ ἡδονὴν αὐτὴν οὐδεὶς ἄν ποτε προσείποι ψευδῆ). (37e10–38a2)

This is a crucial passage—mainly because Socrates, seemingly deliberately, commits a fatal mistake by weakening his position and giving up his defence of hedonic cognitivism. Whereas he was arguing

---

metaphor is used in the *Theaetetus* (194a2–4) to explain how false belief is possible: ‘Like a bad archer [the person who has a false belief] shoots beside the mark and misses it (παραλλάξαι τοῦ σκοποῦ καὶ ἀμαρτεῖν) and it is just this which is called error or deception (ὁ δὲ καὶ ψεῦδος ἄρα ὠνόμασται).’

<sup>273</sup> Cf. 37e10–38a2 where Socrates equates truth and correctness and falsity and incorrectness.

that pleasures are roughly similar to beliefs in the sense that both attitudes have content in virtue of which they can be either true or false, he suddenly starts defending the wholly different idea that pleasures often ‘arise in us (ἡμῖν γίγνεσθαι, 37e11)’ or ‘come to us (ἐκάστοις ἡμῶν ἐγγίγνεσθαι, 38a8)’ ‘accompanied by’ or ‘going with’ (false) beliefs (37e10–38a8). On this weaker picture, the representational attitudes or elements do not constitute the pleasures, they merely serve as some kind of cognitive base or foundation of the pleasures at hand—and this means that the pleasures that are accompanied by false beliefs can, *at best*, be derivatively false.<sup>274</sup>

Protarchus readily snatches the clue and immediately turns the tables on Socrates whose observation was indeed spot on. Even if those cases in which a pleasure is accompanied by a belief, the pleasure we are dealing with remains immune to error—it is the associated belief that may be false. This, then, is the second iteration of Protarchus’ Immunity to Error Thesis:

In so-called cases of ‘false pleasure,’ we are in fact dealing with a pleasure that is merely accompanied by a false belief. Put differently, it is the associated belief that is false, not the pleasure itself.

An example can illustrate the point Protarchus is trying to make. Imagine that the immense gustatory pleasure I am experiencing while drinking a glass of red wine is accompanied by, or parasitic on, the false belief that the wine is a fancy *grand cru classé* although,

---

<sup>274</sup> We should wonder why Socrates ruins his own analogy and does not argue that, at least in a *derivative* sense, we can maintain a distinction between ‘true’ and ‘false’ pleasures. I find myself in agreement with Frede (1985: 167) who astutely observes that ‘if Socrates undermines his own analogy, this must indicate that he himself is not satisfied with it but wants to establish the truth or falsity of pleasure in the primary [non-derivative] sense.’

unbeknownst to me, I am actually drinking a plain *vin de table*.<sup>275</sup> On Protarchus's theory, this does not detract from the truth of my pleasurable experience: it is the connected δόξα—the mistaken assumption about the wine I am drinking—that is false, the pleasure itself is immune to error. This is so, as I have suggested above, because Protarchus denies that there is an appearance-reality gap in the hedonic domain: pleasure is a bit of felt experience and, along with pain, 'just what it is.'

#### 2.4. The Second Hedonic Cognitivism Argument (38b6–41a4)

To counter Protarchus 'spirited defense of pleasure (προθύμως ἀμύνεις τῷ τῆς ἡδονῆς λόγῳ τὰ νῦν, 38a3–4),' Socrates resuscitates his stronger position—hedonic cognitivism—by finetuning and fleshing out his earlier line of argumentation. One of the primary goals of this second stretch of argumentation, I submit, is to convince Protarchus of the fact that pleasure cannot be identified or understood in a way independent of cognitive or representational attitudes. Put differently, Socrates aims to explode the weaker, more deflationary position, previously toyed with by himself, which suggests that in alleged cases of false pleasure we are in fact dealing with false accompanying beliefs. It is not just the case that pleasure is often associated with (false) belief, the idea at play is way stronger: this argument proposes that the two are *merged* and that our pleasures are—at least partially—*constituted by* cognitive or representational states.<sup>276</sup>

---

<sup>275</sup> Bloom (2010: 45).

<sup>276</sup> Cf. Frede (1985: 166 n32): '[A] merger seems to be what Plato has in mind' and Frede (1997: 247): 'Statt sich mit möglicherweise anfechtbaren Parallelen zwischen Meinung

There are a couple of ways in which this second line of argumentation improves upon Socrates's earlier argument for fallibilism. Whereas the rather abstract and hasty First Hedonic Cognitivism Argument more or less stipulated that pleasures have content, leaving it unexplained where this content comes from and which mental operations are involved in the process, the Second Hedonic Cognitivism Argument is more elaborate and more detailed. Having examined the psychological operations driving our affective experiences, Socrates suggests that our experiences of pleasure have representational or cognitive content because they centrally involve representations of (obtaining or non-obtaining) states of affairs, based on (true or false) beliefs.<sup>277</sup> This more fine-grained, more sophisticated picture should make it obvious that, and how, a pleasure can be false, or true, in the literal, representational, semantic sense of the word: false if its constitutive representational states misrepresent the world, true if they represent things correctly.

There is a second way in which this argument improves upon Socrates's earlier attempt to draw on cognitivism to establish hedonic fallibilism. In contrast to that earlier argument, the focus of the current argument lies on *anticipatory pleasure* where it is obvious, or at least *more* obvious, that representational elements and interpretations play a crucial role.

The argument starts with the question of how the human soul or mind comes by its δόξαι which are sometimes false and sometimes true. To account for the truth and falsity of our beliefs, Socrates's basic move is to compare the human mind to a book in which a writer (γραμματεύς) is at work. When engaging in a dialogue with another

---

und Lust zu begnügen, will er vielmehr zeigen, daß gewisse Arten von Lust und Unlust deshalb im eigentlichen Sinne wahr oder falsch sein können, weil sie selbst in Meinungen oder Annahmen über Sachverhalte *bestehen*.' (emphasis in original)

<sup>277</sup> Cf. Szaif (2021: 136).

person or with ourselves—that is, when *thinking*—our beliefs turn into λόγοι (assertions, statements, or propositions).<sup>278</sup> These public or private articulations of our beliefs can either be true or false, depending on whether they describe reality correctly: ‘If what is written is true (ἀληθῆ), then we form a true belief (δόξα ἀληθῆς) and a true account (λόγοι ἀληθεῖς) of the matter, but if what our scribe writes is false (ψευδῆ), then the result will be the opposite of truth (τάναντία τοῖς ἀληθέσιν, 39a4–7).’

It turns out, in addition, that besides the writer there’s another craftsman at work in our mind—a painter or illustrator (ζώγραφος). This painter follows the writer and turns the writer’s words into images or illustrations (εἰκόνας or φαντάσματα)—representations, that is—which are also capable of being true or false, depending on whether they are in line with the beliefs they aim to depict: ‘Are not the pictures of true judgments and assertions true, and the pictures of the false ones false (αἱ μὲν τῶν ἀληθῶν δοξῶν καὶ λόγων εἰκόνες ἀληθεῖς, αἱ δὲ τῶν ψευδῶν ψευδεῖς, 39c4–5)?’

A pivotal question is, of course, how we should unpack this elaborate metaphor of the book and the craftsmen. Frustratingly, Socrates is not very clear on this score. It seems textually plausible and philosophically attractive, though, to assume that the writer at work in the human mind stands for our memory, whereas the painter stands for φαντασία—here to be understood as the psychological faculty that is responsible for the creation of φαντάσματα (representations or mental images).<sup>279</sup>

---

<sup>278</sup> As Lorenz (2006: 105n21) notes, ‘the word λόγος does double duty here ... : it denotes, first, the mind’s articulation of its experience; second, the person’s utterance of a belief in speech.’

<sup>279</sup> Cf. e.g. Delcomminette (2003), Frede (1985), (1993), (1997), and Lorenz (2006); for an overview of imagination in Plato, see Silverman (1991). In a recent paper, Fletcher (2021) argues instead that these φαντάσματα stand for our pleasures.

Let me illustrate this with an example. Suppose I am wandering through a forest when suddenly my mind is bombarded by a bundle of visual and auditory sense data of a feathery, large, and screeching entity that is sitting on a branch, flapping its wings. Suppose, further, that as a consequence of this sensory input I come to have the following δόξα: ‘That thing over there is an owl.’ In line with Socrates’s metaphor, my memory goes to work as soon as I start talking or thinking—that is, talking with myself—about the belief at hand. It does so by storing the belief in the form of an assertion, statement, or word (λόγος) somewhere in my memory.<sup>280</sup>

If it were not for the writer’s memorizing activities, all of our beliefs would be fleeting states, disappearing as soon as the perceptions responsible for them disappeared, and we would have no network of short- and long-term beliefs about ourselves and the world.<sup>281</sup> Stronger still, further down the line Socrates seems to let go of the idea that our memory stores already existing beliefs. Instead, he seems to believe that our beliefs are formed or come to be (ἐν ἡμῖν γιγνόμενοι) when our memory is writing in our minds: ‘When [it] writes what is true (ἀληθῆ), then we form a true belief (δόξα ἀληθῆς) and true account (λόγοι ἀληθεῖς) of the matter, but if what our scribe writes is false (ψευδῆ), then the result will be the opposite of truth

---

<sup>280</sup> As the earlier discussion at 34a10b–9 suggests, memory (μνήμη) must be distinguished from recollection (ἀνάμνησις): while memory preserves our perception and thus stores things, recollection as it were *restores* our memories and takes place ‘when the soul recalls as much as possible by itself what she had once experienced together with the body.’

<sup>281</sup> Cf. the earlier definition at 34a10 of memory as ‘[the] preservation of perception’ (σωτηρίαν ἀσθησέως); cf. Frede (1993: 42, n1): ‘The truth and falsity are initially those of simple sense-perceptions. But the scribe in our soul does more than that; he inscribes long-term views about the world [and] these are true or false in the same sense as the simple statement, however. They depict the way the world is, or they fail to do so.’



(τὰναντία τοῖς ἀληθείσιν, 39a3–7).’

At the same time (ἐν τῷ τότε χρόνῳ, 39b4), the painter or φαντασία jumps in and begins to draw mental images (εἰκόνας or φαντάσματα) of the beliefs I have and the λόγοι that have just been stored in my memory. Whereas my memory arguably stores my beliefs in the form of abstract linguistic units, my imagination is able to make these lifeless linguistic units visible by turning them into vivid mental images. For as Socrates himself explains, the painter is at work ‘when a person takes his beliefs and statements (τὰ δοξαζόμενα καὶ λεγόμενα) from sight or any other sense-perception (τινος ἄλλης αἰσθήσεως) and then somehow views (ὁρᾷ πως) the images of those beliefs and assertions (τὰς τῶν δοξασθέντων καὶ λεχθέντων εἰκόνας) inside himself (ἐν αὐτῷ, 39b9–c2).’

Put succinctly, then, the writer in my head records a belief by turning it into a λόγος and then hands it over to the painter whose job it is to depict its content as an εἰκὼν or φάντασμα.<sup>282</sup> Together, the craftsmen account for the fact that our inner worlds consist of large networks of both linguistic and pictorial units that are intimately connected with our beliefs about the world. That is to say, we have beliefs about the world which are accompanied by vivid mental images putting certain states of affairs lively before our mental, introspective eyes. Socrates adds, furthermore, that this psychology process of belief formation and representation creation is not confined to the past and the present: we also form δόξαι and related λόγοι and εἰκόνες or φαντάσματα about future states of affairs.

Before we turn to Socrates’s final move in the Second Hedonic Cognitivism Argument, there are two important things that bear

---

<sup>282</sup> Note that these images should not be considered exclusively visual: Socrates makes it clear, after all, that we can form images of all kinds of perceptions (39a1–7). Since we have no reason to think that the εἰκὼν or φάντασμα of a sound is visually stored, we should interpret the illustrator metaphorically.

mentioning. To begin with, note that our beliefs and their articulations (λόγοι) are true or false if they depict the world the way it is or if they fail to do. Something similar holds for our representations: they are true if they depict (articulations of) true beliefs, they are false if they depict false beliefs and λόγοι. So if my belief and the connected λόγος about the owl turn out to fail to correspond to reality, they will become false themselves and the same holds for the mental images I have created on the basis of this false belief and the false λόγος attached to it. Truth, then, is at all levels a matter of correspondence—correspondence, that is, between the beliefs, their articulation in λόγοι, and the illustrations in our minds and the world out there. In other words, truth is a matter of correct representation whereas falsity boils down to *misrepresentation*.

With this psychological groundwork in place, the argument renews the earlier defence of the possibility of hedonic mistake and false pleasure. As a first step, Socrates reminds Protarchus of the existence of ‘anticipatory pleasures and anticipatory pains about future states of affairs (τὸ προχαίρειν τε καὶ προλυπεῖσθαι περὶ τὸν μέλλοντα χρόνον, 39d3–4),’ something on which they had already agreed in an earlier stage of the dialogue. Sometimes we feel pleasure when picturing or thinking of something pleasurable that lies ahead of us (e.g. daydreaming about a journey around the world that I plan to undertake next year) and sometimes we feel pain when picturing or thinking of something painful that lies ahead of us (e.g. worrying about a root canal treatment that will be carried out tomorrow).

It is important to notice *why* Socrates uses precisely these anticipatory, future-oriented pleasures to establish the possible (literal) falsity of pleasures. He states, after all, that his theory is omnitemporal: both past, present, and future pleasures can be false.<sup>283</sup>

---

<sup>283</sup> E.g. at 40d7–10.

Note, though, that in the case of pleasures about the past or present, it would have been much more difficult for Socrates, if not impossible, to convince Protarchus of the fact that pleasures have content in virtue of which they can be false. In the case of future-oriented pleasure, as Frede sharply notes, there is often ‘a wide discrepancy between what is enjoyed “as a fact” and what is in fact going to happen’ and this gap creates a lot of room for misrepresentation and falsity. Indeed, in the case of future-oriented pleasure we *only* have our representation of what the future is going to look like. As Frede sums it up, ‘what we enjoy is the anticipated event as it is described or pictured in our minds itself.’<sup>284</sup>

What this suggests, then, is that if Socrates had focused on pleasures in the present, he might have obscured the fact that what we tend to enjoy is not the *thing* itself but the thing as *conceived of* by us. The thought here, to borrow a phrase from Anscombe, is that we always enjoy something *under a description*—especially when it comes to our anticipatory pleasures which necessarily involve some kind of interpretation of, or hypothesis about, the way in which the future is going to unfold.<sup>285</sup>

The second step of the argument (39d7–39e7) is to stress, again, that people’s minds contain writings and pictures about the future. These are called hopes or expectations (ἐλπίδας) and we human beings are said to be ‘forever brimful of expectations, throughout our lifetime (ἡμεῖς διὰ παντὸς τοῦ βίου ἀεὶ γέμομεν ἐλπίδων, 39e5–6).’<sup>286</sup>

---

<sup>284</sup> Frede (1985: 174–5); cf. Frede (1993: xlvii).

<sup>285</sup> Cf. Carone (2000: 275).

<sup>286</sup> Frede and Gosling’s rendering of ἐλπίδας as ‘hopes’ cannot be correct for the simple reason that Socrates explicitly talks about future *pains* as well (e.g. at 39d4 and 40c6.) Since we can and do indeed *expect* pains but do not *hope* for them, I render ἐλπίδας as the more neutral ‘expectations’ (which does not strain the Greek, cf. LSJ s.v.). Szaif (2021: 53n7) raises a different but similarly important point: ‘One can

It is, so to say, a fundamental characteristic of the human condition that we create and entertain beliefs and other mental representations about what lies ahead of us, although we cannot access the future in an epistemically reliable way.<sup>287</sup>

In a third and rather surprising step (39e8–40a2), the argument introduces an ethical distinction between a good man and a bad man.<sup>288</sup> The good man is ‘just, pious, and good in all respects (δίκαιος ἀνὴρ καὶ εὐσεβὴς καὶ ἀγαθὸς πάντως)’ and, as such, ‘also loved by the gods (θεοφιλῆς, 39e10–11).’ The bad man, by contrast, is ‘unjust and in all respects evil (ἄδικός τε καὶ παντάπασι κακός, 39e13)’ and, as such, ‘[the good] man’s opposite (τὸναντίον ἐκείνῳ, 40a1),’ that is, hated by the gods.

The fourth step (40a3–40b1), lastly, ties together the preceding threads and aims to establish the possibility of false pleasures by showing how the representational states that are involved in them are false. Having argued that there are λόγοι—articulated beliefs, that is—and especially (καὶ δὴ καὶ) representations or ‘painted images (τὰ φαντάσματα ἐζωγραφημένα, 40a9)’ in us that we call expectations, Socrates offers an example of a person who ‘sees himself getting an enormous amount of gold and a lot of pleasures as a consequence (ἐαυτῷ χρυσὸν γιγνόμενον ἄφθονον καὶ ἐπ’ αὐτῷ πολλὰς ἡδονάς, 40a10–12)’ adding that it is ‘indeed central to this inner picture that he is beside himself with delight (καὶ δὴ καὶ ἐνεζωγραφημένον αὐτὸν

---

knowingly hope against the odds, while expectation entails that one view the outcome as probable.’

<sup>287</sup> As the chorus in Sophocles’ *Oedipus King* (489) sings, it is characteristic of the tragic human predicament that ‘we fear, but cannot see, what is before us.’

<sup>288</sup> Kenny (1960) was the first to heavily stress this step. See Warren (2003: 129–156) for a thorough examination of the connection between (ethical) character and false pleasure in this argument; Carpenter (2006) also tries to make philosophical sense of this part of the argument without seeing it as ‘the straightforward nod to conventional piety.’

ἐφ' αὐτῷ χαίροντα σφόδρα καθορᾶ, 40a9–12).'

Without unpacking or explaining his own example of the man enjoying the prospect of obtaining a huge amount of gold, Socrates immediately rushes to the conclusion of his argument in the fifth and final step (40b2–e5). It runs as follows:

[T3.4] [Socrates] Now, do we want to say that in the case of good people these pictures are usually true because they are loved by the gods (τοῖς μὲν ἀγαθοῖς ὡς τὸ πολὺ τὰ γεγραμμένα παρατίθεσθαι ἀληθῆ διὰ τὸ θεοφιλεῖς εἶναι) while quite the opposite usually holds in the case of bad people, or is this not what we ought to say?

[Protarchus] That is just what we ought to say.

[S.] And bad people nevertheless have pleasures painted in their minds (τοῖς κακοῖς ἡδοναί οὐδὲν ἦττον πάρεισιν ἐζωγραφημέναι), even though they are somehow false (ψευδεῖς δὲ αὐταί που)?

[P.] Right.

[S.] So bad people as a rule enjoy false pleasures (ψευδέσιν ἡδοναῖς τὰ πολλὰ οἱ πονηροὶ χαίρουσιν), but the good among mankind true ones (οἱ ἀγαθοὶ τῶν ἀνθρώπων ἀληθέσιν)?

[P.] Quite necessarily so.

[S.] From what has now been said, it follows that there are false pleasures in human souls that are quite ridiculous imitations of true ones (ψευδεῖς ἐν ταῖς τῶν ἀνθρώπων ψυχαῖς ἡδοναί, μεμιμημέναι μέντοι τὰς ἀληθεῖς ἐπὶ τὰ γελοιότερα), and also such pains. (40b2–c6)

It seems relatively easy to capture the general gist of the argument and to see what explanatory work the example is supposed to do. Recall that pleasure, as the First Hedonic Cognitivism Argument had aimed to establish, is a unitary complex consisting of two inexorably connected parts or aspects: the attitude of taking pleasure and the content of this act of taking pleasure. This second part of one's pleasure is very often a δόξα (articulated, in speech or in thought, in a λόγος recording or articulating the belief) or an εἰκὼν or φάντασμα in its turn depicting the belief at hand. This happens when we enjoy

our anticipations or advance instalments of future enjoyment (προχαίρειν): when right now, for example, I thoroughly enjoy future pleasures painted in my mind, that is, mental representations of me travelling the world, say, swimming in South American waterfalls, riding camels in African deserts, and crossing the Indian Himalayas on a Royal Enfield motor bike—all along ‘being beyond myself with delight.’

When these painted pictures of future pleasures that I am currently enjoying turn out to be *false*, that is, when the content of my current anticipatory pleasure turns out to be false, my anticipatory pleasure as a whole becomes false too. The same goes for my past or concurrent pleasures whose representational content misses the mark and gets things wrong. In that respect pleasure and belief are indeed remarkably similar: both derive their truth and falsity from the semantic or representational value of their content. The hedonic content, to use Socrates’s own neat metaphor, *fills* or *infects* the pleasure as a whole so to speak with its own shaky predicament of falsehood.<sup>289</sup> As Socrates drives home the conclusion of the Fallibilism Argument:

[T3.5] [Socrates] Whoever believes anything at all is always *really* believing (δοξάζειν μὲν ὄντως ἀεὶ τῷ τὸ παράπαν δοξάζοντι), even if is not about anything that is the case—be it in the present, past or future (μὴ ἐπ’ οὐσι δὲ μὴδ’ ἐπὶ γεγονόσι μὴδὲ ἐπ’ ἐσομένοις).<sup>290</sup>

---

<sup>289</sup> As Socrates himself puts it at 42a7–9, when he introduces the second type of hedonic falsity: ‘Earlier it was true and false beliefs which ἀνεπίπλασαν (filled or infected) the respective pleasures and pains with their own condition (τοῦ παρ’ αὐταῖς παθήματος).’ For these different senses of ἀναπίμπλημι, see LSJ (s.v. II and II.2).

<sup>290</sup> Pace Frede (1993: *ad loc.*) and Fowler (1962: *ad loc.*), it makes best sense to take these instances of εἶναι (and those below) veridically rather than existentially: false pleasure

[Protarchus] Right.

[S.] And these were, I think, the conditions that produce a false belief and believing false (τὰ ἀπεργαζόμενα δόξαν ψευδῆ καὶ τὸ ψευδῶς δοξάζειν), weren't they?

[P.] Yes.

[S.] But should we not also grant to pleasures and pains a condition that is analogous in these ways (τὴν τούτων ἀντίστοροφον ἕξιν ἐν ἐκείνοις)?

[P.] In what ways?

[S.] In the sense that whoever has any pleasure at all, however ill-founded it may be (τῷ τὸ παράπαν ὅπως οὖν καὶ εἰκῆ χαίροντι), really does have pleasure (χαίρειν ὄντως ἀεὶ)—even if sometimes it is not about anything that either is the case or ever was the case or often (or perhaps most of the time) refers to anything that will ever be the case (μὴ μέντοι ἐπὶ τοῖς οὔσι μὴδ' ἐπὶ τοῖς γεγονόσιν ... ἐπὶ τοῖς μὴδὲ μέλλουσί ποτε γενήσεσθαι).

[P.] That also must be necessarily so.

[S.] And the same account holds in the case of fear, anger, and everything of that sort (ὁ αὐτὸς λόγος ἂν εἴη περὶ φόβων τε καὶ θυμῶν καὶ πάντων τῶν τοιούτων)—namely that all of them can at times be false (ψευδῆ πάντα τὰ τοιαῦτα ἐνίοτε)?

[P.] Certainly. (40d7–e5)

This bit of text confirms my claim that Socrates's Second Hedonic Cognitivism Argument is an attempt to resuscitate his First Hedonic Cognitivism Argument—and especially its central proposal that, like a belief, a pleasure can be false because it has content. In addition, Socrates makes it clear that hedonic cognitivism is omnitemporal—it applies to *all* pleasures—even though his argument had mainly focused on future-directed, anticipatory pleasure. Stronger yet, it applies to *all* passions or emotions. In addition, Socrates at the same time concedes to Protarchus that he was in a way right to think that there is something incorrigible about our subjective hedonic

---

misrepresents what is the case, what is true, what is a fact, it does not misrepresent 'realities' (Fowler) or 'what exists' (Frede).

experiences: even if a pleasure's content is false, we are still 'really having' such a false or 'ill-founded or fortuitous' pleasure.<sup>291</sup>

### 3. The Falsity of False Pleasure

One vexing and crucial question remains, however.<sup>292</sup> As Socrates puts it, some pleasures are 'somehow ( $\pi\omicron\upsilon$ , 40b7)' false. What exactly renders a pleasure false, though, 'if [the pleasure] is not about anything that either is the case or ever was the case or often (or perhaps most of the time) refers to anything that will ever be the case'? Where, in other words, should we locate the truth or falsity of a pleasure?

#### 3.1 An Interpretive Dichotomy

It is often believed that the most natural reading of Socrates's argument and its central example is a factual or descriptive reading.<sup>293</sup> On such a reading, a person enjoying a pleasure with the content that  $p$  enjoys a false pleasure if  $p$  is not the case.<sup>294</sup> This is the case, to stay

---

<sup>291</sup> Note, though, that Socrates's third type of hedonic falsity goes on to differentiate between merely apparent and genuine pleasure, thus suggesting that those who experience merely apparent pleasure actually go wrong in their self-ascriptions of pleasure: they are hedonically fallible in the strongest sense of the word. For discussion of this argument, see chapter 2 of this dissertation, Pearson (2019), and Whiting (2014).

<sup>292</sup> As Warren (2003: 131) expresses this worry, 'Socrates himself does not make clear precisely where any falsehood enters into the imagined example.'

<sup>293</sup> Delcomminette (2003), Evans (2008-a), Frede (1985), (1992), (1993), and (1997), Hackforth (1945: 72–73), Penner (1970), Williams (1959).

<sup>294</sup> This is even conceded by those who read the argument evaluatively (such as Harte, Moss, Russell).



close to Socrates's example, when a person experiences pleasure while picturing himself obtaining a lot of money next week, although the representational content of his current anticipatory pleasure does not depict the relevant future state of affairs correctly: when next week arrives, say, his lottery ticket turns out not to be the winning ticket.<sup>295</sup>

In these cases, Plato's point is that the content of one's pleasure does not stand in a correct representational relation with reality: it misrepresents some way the world really is. There is a mismatch, in other words, between our representation or depiction of some state of affairs in our life and how that state of affairs actually is, was, or will turn out to be. As a result, all these pleasures have content that is false which renders the pleasures themselves false too—in the same way in which the false content of a belief 'infects' the belief as a whole with its falsity.

Despite the *prima facie* plausibility and naturalness of this reading, recent scholarship has witnessed the increasing popularity of a different, evaluative or ethical reading of the argument. Space does not allow me to go too deeply into the reasons for the introduction of this line of interpretation. Most importantly, Socrates's moralistic digression—in which he ascribes false pleasures to *bad* people—suggests that there must be something evaluatively or ethically sketchy about the false pleasures under discussion.<sup>296</sup> As Frede puts it,

---

<sup>295</sup> Even though his argument focuses on anticipatory, future-directed pleasure, Socrates makes it clear that his analysis is omnitemporal: whether a pleasure is concerned with the past, the present, or the future, it always involves a representation of some way the world was, is, or will be and this creates space for falsity. I might enjoy a painting, say, merely because I think it is an original Van Gogh—even though I am actually looking at a fake, crafted by a master imitator. In a similar vein, the pleasure I enjoy while remembering some childhood event I am particularly fond of might in fact centrally involve a confabulated memory.

<sup>296</sup> It might also be argued that it takes an evaluative reading to make sense of Socrates's claims that we can analyse other passions, such as anger and fear, in broadly

Socrates's emphasis on a putative connection between the truth of one's pleasures and one's moral standing seems to be implying that Plato also holds that '*the moral content* of foolish pleasures is mistaken, so that they represent a skewed view of life.'<sup>297</sup> (my emphasis)

If a person experiences a pleasure with the content that *p*, the evaluative reading suggests that their pleasure is false if it is (or would not be) good (or otherwise positively evaluatively charged) that *p*.<sup>298</sup> The view that pleasure is true or false depending on whether the object or state of affairs in which it is taken is actually valuable is defended, for instance, by Hampton who claims that 'false pleasures result from the inability to grasp what is truly valuable in life.' Thus, according to Hampton, the person in Plato's example—who envisions himself 'getting an enormous amount of gold'—'enjoys a pleasure that is false because it is based on the false belief that money is intrinsically valuable.'<sup>299</sup>

In her defence of the possibility of false pleasure, Lovibond—who is clearly indebted to this stretch of argumentation in the *Philebus*—likewise starts from the claim that there is an 'internal relation between pleasure and value-perception.' Departing from this connection between the two, she goes on to suggest that something counts as a false pleasure just in case we 'refuse to accept the construction of our experience of the bad object as a pleasurable experience.'<sup>300</sup>

---

similar ways (40e2–5) and the puzzling claim that it is impossible 'to account for the badness of pleasures (πονηραὶ) unless they are false (πλὴν τῶ ψευδεῖς εἶναι, 40e6–8).')

<sup>297</sup> Frede (1993: 44); cf. Marechal (2021: 285). For broadly similar views, see Carpenter (2006: 18–19), Hackforth (1945:73), Harte (2004: 125), Russell (2005: 181–2), Warren (2014: 136–146).

<sup>298</sup> Russell (2005: 176, 182).

<sup>299</sup> Hampton (1987: 255).

<sup>300</sup> Lovibond (1989: 222).

According to Dimas, the *Philebus* aims to offer an account of ‘good and bad pleasures’ whose key claim is that ‘a pleasure is true if it is taken in what is good—and false if it is taken in what is not good—for the agent’<sup>301</sup> and Marechal’s recent interpretation concurs: a pleasure is true ‘when it correctly represents the actual value of things.’<sup>302</sup> According to Moss, finally, the *Philebus* defends the following thesis: ‘to feel a passionate pleasure in  $x$  is to hold that  $x$  is good (or in some more determinate way valuable), and thus such a pleasure will be false just in case  $x$  is not good, true just in case  $x$  is good.’<sup>303</sup> (emphasis in original)

The most careful, most sustained articulation of this type of interpretation, however, can be found in Russell’s *Plato on Pleasure and the Good Life*. According to Russell, it is not enough to appreciate that Plato thinks pleasures are *about* something—that they have intentional content, that is—what is even more important is to see that, for Plato, our pleasures ‘represent their objects to the agent under such descriptions as ‘satisfying’, ‘worth-while’, ‘just what I need’, and so on, and that, as such, they can also misrepresent their objects.’<sup>304</sup>

Crucial to Russell’s evaluative reading is a specific understanding of the nature of φαντασία and the role it plays in how we apprehend the world. Rather than merely representing the world as it is, or matter-of-factly, φαντασία represents the world in a concern-laden way—that is, the faculty of φαντασία brings the *significance* of certain states of affairs for a person into vivid relief. In Russell’s words:

---

<sup>301</sup> Dimas (2019).

<sup>302</sup> Marechal (2021: 283).

<sup>303</sup> Moss (2012: 15).

<sup>304</sup> Russell (2005: 176–177).

The paintings do not merely tell us more than the writings do. It is by 'painting' them within ourselves that we overlay our experience, so to speak, with the vivid shape and color that our concerns lend. Pleasure goes beyond taking things to be such and such, to ascribing them a value and power that our deeper concerns invest in them.<sup>305</sup>

Although Russell does not flesh out or defend this understanding of φαντασία in much detail, let us try to make better sense of it.<sup>306</sup> In Plato's *Sophist*, the Eleatic stranger defines φαντασία as 'the δόξα which arises not on its own'—as part of an internal, soundless dialogue the mind has with itself—'but through perception (μη καθ' αὐτὸ ἀλλὰ δι' αἰσθήσεως, 264a4–6)' or, alternatively, as 'a mix or commingling of perception and δόξα (σύμμειξις αἰσθήσεως καὶ δόξης, 264b2).' To see what is meant by this, let us take the example Socrates gives at the start of his argument (at 38c2–d11). The example concerns a person who is looking at figure from a distance, wonders what it is he is seeing, and then represents it to himself (φανταζόμενον; cf. φαντασθέντα) as a man, although it is actually a statue.<sup>307</sup>

In this situation, the man's perception is as it were structured by a δόξα or, to put the same point slightly differently, his perception is subsumed under a concept ('man'). That is to say, the raw perceptual input is mistakenly interpreted or represented to him as a man rather than as a statue—as it should have been.<sup>308</sup> It might very

---

<sup>305</sup> Russell (2005: 179).

<sup>306</sup> In my attempt to clarify why Russell understands φαντασία the way he understands it, I am leaning heavily on the analyses we find in Lorenz (2006: 105–110) and Delcominette (2003: 222–225).

<sup>307</sup> The notion of error and deception already figures here: if this person represents the figure as a man, he gets things right (ἐπιτυχῶς), but if he represents the figure as a statue he is mistaken or gets misled (παρενεχθείς).

<sup>308</sup> Cf. Lorenz (2006: 105): 'In the case that Socrates is mainly interested in, someone forms a false perceptual belief, misidentifying a man in the distance as a statue (38c5–

well be possible, therefore, that one's perception is strictly speaking identical in two cases, although the way in which things are represented to someone differs from case to case—depending on the δόξα that shapes or structures the perceptual input at hand.<sup>309</sup>

On the assumption that φαντασία is defined by Plato as a commingling of perception and belief or a belief which arises through perception, it follows that our appearance, or the product of φαντασία, can be false for two reasons. It can be false because something goes wrong with the perceptual part of the appearance and it can be false because something goes wrong with the cognitive part of the appearance or the belief we impose upon the perceptual input.

On the first construal, my perception might be blurred or vague—I am not seeing clearly because I am looking at things from a distance, as Socrates suggests—so that I accidentally apply the δόξα 'that is a man' to my perception. In this case, I am struck by the appearance *as of* seeing a man, although I am actually looking at a statue. Alternatively, the concept that I apply to the perceptual input might *itself* be defective. I might, for instance, conflate real men and fake statues because my grasp of the notions 'man' and 'statue' is shaky.

From this latter case of falsehood, it is clearly a small step to an evaluative understanding of Socrates's argument. Although it might

---

e7). ... In this case, the painter's work depicts, not what the person in fact saw (a man), but what he falsely believes he saw (a statue). As Socrates says, the painter follows the scribe, and what the painter paints is true or false depending on what the scribe writes (39c4–5, with 39a). The painter's works thus involves interpretation of what one saw or perceived otherwise. They depend on the rational states or dispositions that are one's perceptual beliefs.'

<sup>309</sup> Interestingly, the *Timaeus* (28a2–3, 52a7) holds that it is only by means of this combination of perception and δόξα that we gain access to the world, not by means of bare, unstructured perception in and of itself. See McDowell (1994) for a famous defence of the claim that perceptual experiences have conceptual content.

be hard to see how we can be mistaken in our very concepts of ‘man’ and ‘statue,’ it is easier to see how this works in the case of the evaluative δόξαι giving rise to evaluative appearances which figure centrally in our pleasures (such as ‘winning the lottery will be extremely rewarding’ or ‘a new MacBook is just what I need.’) When it comes to these evaluative representations, there is clearly room for conflating concepts of value, goodness, or badness and wrongly taking to be good or valuable what is actually bad or invaluable and *vice versa*.

With this evaluative understanding of φαντασία in place, a fairly sophisticated evaluative reading of the Fallibilism Argument presents itself. Russell offers the instructive example of a rich man who has decided to buy a Jaguar in the near future. Not only does this person have the factual belief ‘I will own a Jaguar soon’, which just dryly picks out a future state of affairs, his φαντασία faculty overlays this future state of affairs with significance and meaning. Like a belief, φαντασία represents the world—but it does so in a more concern-laden way. For instance, the painter inside him depicts his belief about soon possessing a Jaguar with all kinds of positive evaluative images and colors such as ‘the thrill of speeding along, being the envy of my neighbors and friends, [and] increasing my sex appeal’<sup>310</sup> and this highly evaluatively charged painting functions itself, importantly enough, as the content of anticipatory, future-directed pleasure.

The crux, of course, is that such a highly charged evaluative representation of some future state of affairs can get things right or wrong and be false or true. It is true when the future state of affairs does indeed have the significance someone attributes to it, it is false when it does not. Perhaps the Jaguar does not get the man any female attention; perhaps the ever-growing pile of speeding penalties starts

---

<sup>310</sup> Russell (2005: 179).

to bother him; and perhaps his friends are not impressed but hate his bragging about his fancy car. In these cases, his anticipatory pleasures were clearly false. As Russell concludes:

[O]n Plato's analysis there *is* something for pleasure to be right or wrong about. ... For it is all too possible that a person may represent something to herself as satisfying, rewarding, and worthwhile, when, in fact, it is none of those things.<sup>311</sup>

### 3.2. Resisting the Dichotomy: In Defence of a Wider Reading

In sum, then, the descriptive reading of Plato's Fallibilism Argument suggests that I experience a false pleasure if I take pleasure in a descriptive state of affairs  $p$ , although  $p$  does not obtain, whereas the evaluative reading suggests that my pleasure is false if I take pleasure in  $p$  being  $F$ — $F$  being some further determinable positively charged evaluative term—although  $p$  is not  $F$ .

I resist this stark dichotomy for three reasons. Firstly, it strikes me that descriptive and the evaluative readings share an important common ground that is often overlooked or obscured in the literature. In both cases, I argue, false pleasure is a matter of *misrepresentation*. Whenever I experience false pleasure, no matter whether we construe this falsity descriptively or evaluatively, there is mismatch between my representation of some way the world is and how the world actually is: in one case my pleasure misrepresents a purely 'factual' matter, while in the other case it misrepresents an 'evaluative' matter.

This, however, is something that is often missed in discussions about this exegetical dispute. Evans, for instance, ascribes to the descriptive camp the idea that 'what makes a pleasure false is either

---

<sup>311</sup> Russell (2005: 182) (emphasis in original).

its having false content or its sharing its content with some false belief.<sup>312</sup> The evaluative reading suggests, in sharp contrast, that Plato's argument claims that 'what makes a pleasure false is its having content that is *bad*, or at least *not good*.'<sup>313</sup> Evans goes on to give two examples:

[On the descriptive or factual interpretation,] if you are pleased (and believe) that the war is ending, then your pleasure is false if it is not the case that the war is ending. [On the evaluative interpretation,] if you are pleased that the war is ending, then on this view *your pleasure is false if the war's ending would be a bad thing, or would not be a good thing*.<sup>314</sup>

He criticises the evaluative reading of the argument *inter alia* on the grounds that this option carries the prohibitive cost of making the Fallibilism Argument invalid: Socrates would, after all, commit the fallacy of equivocation by shifting thoughtlessly between 'false' in a literal sense and 'false' in some mysterious moral or evaluative sense.<sup>315</sup>

*Pace* Evans, it strikes me that one may simply maintain that according to both the descriptive and the evaluative interpretation of Plato's Fallibilism Argument, falsity is a matter of *misrepresentation* or a *failure of correspondence* between the representational content of one's pleasure and some reality external to that pleasure with which it aims to make contact. To see why this is the case, consider the following instances of false pleasure. For starters, I might be thrilled that I am

---

<sup>312</sup> Evans (2008-a: 92) (my emphasis).

<sup>313</sup> Evans (2008-a: 92–3) (emphasis in original).

<sup>314</sup> Evans (2008-a: 92–93).

<sup>315</sup> Cf. Evans (2008-a: 102) ('[Socrates] gives no indication whatsoever that he expects these propositions to be assessed according to some non-veridical, evaluative standard') and his reconstruction of the argument on p. 98–101. Cf. Shorey (1903: 149n147): 'The epithet false is used as a mere rhetorical expression of the disdain or pity of the overlooker.'



running into an old friend I had not seen for a long while, even though he turns out to be a lookalike stranger. Likewise, I might remember a holiday to a Caribbean Island with great retrospective joy, even though I was in fact bored for the majority of the time while lying on the beach.

In a slightly different vein, a young student might experience great anticipatory pleasure before embarking upon a PhD trajectory because he represents the life of the mind as satisfying, rewarding, and perfectly aligned with his cares and concerns, although he soon discovers that academia has many powerful negative undercurrents which turns it into a place where you can easily lose yourself. Similarly, a technology lover might enjoy her new iPhone as pleasant ‘in its own right (αὐτὸ καθ’ αὐτό)’ or ‘fine (καλόν),’ as Marechal has suggested in a recent paper, even though this pleasure misrepresents the actual value of things because, like all material possessions, iPhones are (at best) instrumentally or extrinsically valuable.<sup>316</sup>

What these examples should make clear, then, is that the evaluative reading is pretty similar to the descriptive reading—or at least far more similar than it is typically made out to be. In all these cases, whether they are strictly speaking factual or evaluative, a pleasure is false just in case there is a mismatch between the representational content of this pleasure—my interpretation of the world beyond my head—and how things really are. No matter how we gloss these cases of false pleasure, the falsity of any misleading pleasure always consists in a mismatch or failure of correspondence between the cognitive or representational elements we entertain in our heads, and which constitute our pleasures, and some (factual or evaluative) state of affairs in the world beyond our heads. A pleasure is not false, then, just because it is bad; a pleasure is false because its

---

<sup>316</sup> Marechal (2021: 296–299).

representational content fails to latch onto some state of affairs in actual reality.<sup>317</sup>

There is a second reason to resist this dichotomy. Since our choice of interpretation seems left underdetermined by the available textual data,<sup>318</sup> which explains why Socrates's example has been interpreted in various competing ways, charity arguably requires us to opt for the broader reading I have defended here. It suggests that our pleasures can be false in, at least, two different ways and this allows Plato to cast his net wider and argue for a larger set of cases where our pleasures get things wrong.

Thirdly and most importantly, though, there is good reason to believe that, like other emotions, any pleasure necessarily involves both a descriptive and an evaluative cognition: the factual belief that some state of affairs obtains and the evaluative belief—sometimes called 'appraisal' or 'construal' in the psychological literature—that this state of affairs is positively evaluatively charged because it somehow bears positively on the subject's cares and concerns.<sup>319</sup>

An insightful Medieval line of reasoning, recently rehabilitated by Kenny,<sup>320</sup> suggests that an emotion is not just concerned with some object or state of affairs in the world (its target or 'material object'), it also has a 'formal object.' This formal object can be understood as the

---

<sup>317</sup> Although he focuses on anticipatory pleasures, Warren (2013: 145–146) comes close to arriving at the same conclusion: 'the error involved [in Socrates's example] stems from some inconsistency between the present depiction of some future experience ... and how that future experience turns out to be.'

<sup>318</sup> See Warren (2013: 131).

<sup>319</sup> Alternatively, we might be dealing with just one factual-cum-evaluative belief—e.g. the belief that something *F* (positively evaluative charged) is present which contains a factual belief ('there is some *x*') and an evaluative belief ('*x* is *F*'). This seems to be the position of the (Chryssiopian) Stoics who defined pleasure as 'the fresh belief that something good is present.'

<sup>320</sup> Kenny (1963: 187–194).

property I implicitly ascribe to the object or state of affairs in virtue of the fact that I experience the relevant emotion about the target of my emotion.<sup>321</sup> Indeed, if the object or state of affairs did not have a certain character—or if I did not *represent* or *interpret* or *see* it as having that character—I would simply not be able to experience the relevant emotion.<sup>322</sup>

If I am afraid of a snake, for instance, I necessarily consider it as *dangerous* or as possibly *harmful*.<sup>323</sup> Likewise, if I feel angry because of what you did to me, I necessarily consider your action offensive and I cannot feel love towards you, unless I evaluate you as a worthwhile person. In a similar way, I argue, the thing I enjoy must have a certain character or I must at least see it under this description. Whenever I experience pleasure, I necessarily respond to something *as* (represented as) something satisfying—otherwise I would not experience pleasure but remain indifferent. The point, then, is that whenever I undergo an affective or emotional experience, I do not just believe that some object is present or that some state of affairs obtains, I also believe that the target of my experience has a certain character (that bears positively or negatively on my cares and concerns).

That our emotional and affective experiences necessarily

---

<sup>321</sup> Scarantino and de Sousa (2018: §4), de Sousa (2018: §3), Deigh (1994: 834–835). More generally, as de Sousa (2018: §3) points out, for any intentional state, its formal object is ‘that characteristic that must belong to something if it is to be possible for the state to relate to it.’

<sup>322</sup> The thought here, as Deigh (1994: 835) puts it, is that ‘there is a logic to the concept of [emotion] *x* such that to say that a person feels *x* toward *z* implies that he believes such and such about *z*.’

<sup>323</sup> As Socrates defines fear in the *Protagoras* (358d5–6), it is ‘something I describe as an expectation [literally ‘advance belief’] of something evil (προσδοκίαν τινὰ λέγω κακοῦ τοῦτο). Cf. *Laches* 198b8–9: ‘fear is produced not by evils which have happened or are happening but by those which are anticipated (τῶν κακῶν ... τὰ προσδοκώμενα) because fear is an expectation [literally ‘advance belief’] of a future evil (προσδοκίαν μέλλοντος κακοῦ).’

involve a factual and evaluative cognition is corroborated by the observation that these states can be punctured in two ways. The fear I experience while seeing a snake slithering across my path, to use a well-worn Buddhist example, collapses when I discover that I am mistaken about the material object (it is not a snake but a piece of rope) or when I discover that I am mistaken about the formal object (the snake is actually harmless). The same goes for my anger towards you: it will dampen if I discover that it was not you but someone else who wronged me or if I come to the realization that your action was not offensive or otherwise morally problematic. When this happens, the target or focus of my emotion fails to fit its formal object.<sup>324</sup>

In precisely the same way, I want to suggest, the pleasure I experience while thinking I have won the lottery can also collapse in two ways: when I discover that I misread the number of the winning lottery ticket, say, or when I discover that winning a lottery is strongly correlated with depression, anxiety, and suicide and thus not valuable, rewarding, or satisfying.

#### 4. Hedonic Cognitivism, Deceptive Pleasure, and the Good Life

Although Plato is well known for his restorative or improvement indicator model of pleasure, we have just seen that the *Philebus* also develops a cognitivist theory of pleasure on which (part of) its defence of hedonic fallibilism is predicated.<sup>325</sup> On this account, a pleasure is

---

<sup>324</sup> De Sousa (2018: §3).

<sup>325</sup> This raises the interesting question of how we should harmonize these seemingly conflicting restorative and cognitivist accounts. Following Aristotle (*On the Soul* 1.1, 403a–b), it could be argued that they are just two different ways—one philosophical, the other scientific—of describing one and the same thing: ‘The dialectician will define [anger] as desire for retaliation, or something of this sort, while the natural scientist

more than a subjective and private sensation or feeling or mere ‘smooth movement [of the soul] (λείαν κίνησιν τὴν ἡδονήν)’ (as the Cyrenaic hedonists liked to call them).<sup>326</sup> Instead, they are best understood as psychologically complex representational attitudes which make contact with the world or fail to do so. In short, Plato treats pleasure as some kind of openness to the world: like belief, perception, and other representational states, pleasure is a way of apprehending reality or taking the world to be a certain way.<sup>327</sup> To borrow a Proustian expression made famous by Martha Nussbaum, our hedonic experiences can be described as ‘geographical upheavals of *thought*.’<sup>328</sup>

Plato’s hedonic cognitivism involves two key ideas. Firstly, Plato claims that what we usually treat as one thing—our pleasure—can in fact be broken down in two parts, aspects, or faces: an act (*that* we are enjoying ourselves) and the content of this attitude (*what* we are enjoying) which derives from the representational elements (the δόξαι and φαντάσματα) that constitute a pleasure. When the representational content is false, when it gets things wrong, the pleasure as such gets infected with its falsity—even though Socrates concedes to Protarchus that it is impossible to detract from the fact *that*

---

will define it as boiling of the blood and heat around the heart. Of these, one describes the matter (τὴν ὕλην) and the other the form and the account (τὸ εἶδος καὶ τὸν λόγον). For this is the account of the thing, but it is necessary that it be in matter of this sort if it is to exist.’ In the same way, a house can be described as ‘a shelter capable of guarding against destruction’ or as a collection of ‘stones and bricks and timber.’ For discussion, see Charles (2009).

<sup>326</sup> SSR 4b.5 = Eusebius, *Preparation for the Gospel* 14.18.32.

<sup>327</sup> Plato—rather than the Stoics who defined pleasure as ‘a fresh belief that a good thing is present’ (SVF iii.391)—should be credited, then, with the invention of cognitivism about emotions and pleasures. See Moss (2012: 264n12) and Thein (2012: 121n24) who claims that ‘there is an obvious possibility of taking the *Philebus* to be one of the sources of the Stoic conception of emotions.’

<sup>328</sup> Nussbaum (2001) (my emphasis).

one is experiencing pleasure. That is to say, a false pleasure is still genuinely a pleasure and not something else. Secondly, I have suggested that any pleasure necessarily involves the factual belief that some state of affairs  $p$  obtains and the evaluative belief that  $p$  is  $F$  (positively evaluatively charged and closely connected with the subject's cares and concerns).

Accordingly, Plato recognizes two ways in which our pleasures—and other emotional or affective states like fear or anger—can be false: a pleasure can be false because its content represents the hedonic object as good, although it is indifferent or bad, and a pleasure can be false because its content represents the hedonic object as obtaining, although it is not-obtaining.<sup>329</sup> What this suggests, then, is that hedonic truth or falsity is always a matter of (mis)representation. Like any other representational attitude, a pleasure is true—or correct, or accurate, or hitting its aim—if its content represents the world correctly, it is false—or incorrect, or inaccurate, or missing the mark—if its content represents the world incorrectly.

Although Socrates himself suggests that the topic of false pleasure is 'directly relevant (πρὸς τὰ παρεληλυθότα ... προσήκοντα, 36d6–7)' to the discussion he and Protarchus are having, it is far from clear how hedonic fallibilism, as well as the underlying hedonic cognitivism, are meant to integrate with the dialogue's

---

<sup>329</sup> A similar take on Plato's fallibilism can be found in Damascius (*Lectures on the Philebus* 184 and 186) who also believes that Plato recognizes two ways in which a pleasure can be false: if the hedonic content is 'not present' or if it is 'not good'. What this suggests, then, Damascius continues, is that, like belief, pleasure aims at truth, but, unlike belief, it also aims at the good: 'the love of the ugly, or of what is not beautiful, is false, since it is a striving that seeks the beautiful.' Gosling (1975: 50) also broaches this interpretive option (albeit without endorsing it): 'for any given state of affairs, there is an appropriate reaction, and just as when we make a judgement we are aiming at the right judgement about the thing, so when we react, we are aiming at the right reaction.' See e.g. Evans (2008-a: 120 and 120n40) for brief discussion.

inquiry into what makes for a life worth living.

In this section, I turn to this puzzle and canvass what I take to be the most salient ramifications of this stretch of argumentation. More precisely, it strikes me that Plato's hedonic fallibilism, and the underlying hedonic cognitivism, make trouble for Protarchus' naïve thought that pleasure is self-sufficient.<sup>330</sup> For one, the proposal that pleasures have content in virtue of which they have semantic value is an apt illustration of the more general idea that pleasures cannot be understood independently from cognizance. For another, if our pleasures are in the business of representing the world and if we prefer true pleasures over false pleasures which mislead us about some way the world is, we are left to conclude that there are other values—such as truth—besides pleasure.

One of the unexpected take away messages of the *Philebus* is that there is a tight, if not inseverable link between pleasure and cognition—the two initial candidates for the good. Part of Socrates's strategy is to show that pleasure cannot be understood independently from cognition. Indeed, intellection, cognition, reason, or some other kind of mental operation turns out to be a *necessary condition* for pleasure: although 'the intelligent can be pleasant,' as Butler captures the spirit of the *Philebus* nicely, 'the pleasant is, in all ways, intelligent.'<sup>331</sup>

The elenctic part of the dialogue (20e4–21d5) already hints at this strong connection between pleasure and cognition. If we were

---

<sup>330</sup> Cf. 21a14–b2: '[Soc.] Might you not have some need of knowledge, intelligence, and calculation, or anything else that is related to them? [Prot.] How so? If I had pleasure, I would have everything (πάντα γὰρ ἔχοιμι ἄν που τὸ χαίρειν ἔχων)!'

<sup>331</sup> J. E. Butler (2007: 121). This idea also influences the picture of the good life (as a mixture of pleasure and reason) the dialogue ends up espousing: as J. E. Butler (2007: 120) again puts it, 'if you take any slice of the good life,' including one's pleasures, 'there is no chance that you will end up with an intellect-free slice.'

‘devoid of intelligence (κενόν ὄντα πάσης φρονήσεως)’ and did not possess ‘either reason, memory, knowledge, or true opinion (νοῦν καὶ μνήμην καὶ ἐπιστήμην καὶ δόξαν μὴ κεκτημένος ἀληθῆ, 21b6–7),’ our hedonic life as we know and love it could not exist. A lack of cognitive capacities would not just mean that we could not remember past pleasure and plan for future pleasure, it would even make us ignorant of our occurrent pleasures: like some kind of lowly sea creature, we would not have the self-intimating belief that we are enjoying ourselves while experiencing pleasure.

The *Philebus* also suggests that the mind is necessary for the experience of pleasure *überhaupt*. Although Socrates initially claims that pleasure is identical to the restoration of a previously disrupted state of harmony and balance—‘the return towards its own nature, this general restoration, is pleasure (τὴν δ’ εἰς τὴν αὐτῶν οὐσίαν ὁδόν, ταύτην . . . τὴν ἀναχώρησιν . . . ἡδονήν, 32b2–4)<sup>332</sup>—he is quick to add that such a restorative process only turns into a pleasure if it provokes a kind of ‘upheaval (σεισμὸν)’ in the mind and enters conscious awareness. In a way, then, even raw material pleasures involve some mental operation (αἴσθησις) and there is another type of pleasure—including anticipatory and intellectual pleasures—which is solely mental or psychological.

I believe that the hedonic cognitivism fueling Plato’s hedonic fallibilism serves as a powerful third illustration of this general idea that pleasure and cognizance cannot be pulled apart. Its key proposal, recall, is that our pleasures have content and that the human mind is responsible for this content. Were it not for the activity of the writer and the painter in our minds—the psychological craftsmen who are in the business of claiming that some state of affairs obtains and that this state of affairs is positively tied to the things we care about—we would

---

<sup>332</sup> Cf. 32a8–b4 and 31d1–10.



not be able to experience pleasure.<sup>333</sup>

Hedonic cognitivism also suggests that our pleasures can be true or false. Again, Socrates refrains from spelling out how this defence of hedonic fallibilism is supposed to integrate with their attempt to delineate the good life. ‘One is left to assume,’ as de Sousa puts it, ‘that falsity might detract from the claim of any conditions to be life’s chief good.’<sup>334</sup> Throughout the *Philebus*, and especially in the discussion of the typology of false pleasure, a lot of emphasis is put on the association between pleasure and illusion or deception.<sup>335</sup> Towards the end of their discussion, Protarchus concludes that pleasure is ‘the greatest impostor of all (ἀπάντων ἀλαζονίστατον, 65c5)’ and throughout his account of false pleasure, Socrates repeatedly claims that such pleasure pulls appearance and reality apart.<sup>336</sup> Indeed, the

---

<sup>333</sup> Note that if the cognitive activities of the human mind are responsible for the content of our pleasures, this undermines another idea defended by Protarchus in the opening gambits of the dialogue. When Socrates tried to force him to accept that the pleasure of the moderate or wise person is different from the pleasure of the immoderate or foolish person, Protarchus responds as follows: ‘Well, yes, Socrates—the pleasures come from opposite things (ἀπ’ ἐναντίων), but they are not at all opposed to one another (οὐ μὴν αὐταὶ γε ἀλλήλαις ἐναντίαι). For how could pleasure not be, of all things, most like pleasure (ἡδονῆ γέ ἡδονῆ μὴ οὐχ ὁμοίωτατον ἂν εἶη ... πάντων χρημάτων)? How could that thing not be most like itself (τοῦτο αὐτὸ εἰαυτῶ)?’ (12d7–e2) Plato’s Fallibilism Argument suggests, however, that pleasures are differentiated or individuated on the basis of their content in precisely the same way in which excitement and nervousity are differentiated on the basis of their content despite their identical phenomenal feel. What Plato suggests, in other words, is that the differentiae of specific, individual pleasures are not phenomenal, they are cognitive. See de Sousa (2018: §2) for good discussion of this line of thought.

<sup>334</sup> De Sousa (2013: 125).

<sup>335</sup> Moss (2006: 504, 533n64) and Mooradian (1992).

<sup>336</sup> Cf. e.g. 42b9–1c for the distinction between ‘mere appearance’ and ‘real being (τὸ φαίνόμενον, ἀλλ’ οὐκ ὄν)’ and 44c8–9 where merely apparent (rather than genuine) pleasure is associated with trickery or witchcraft and untrustworthiness or unsoundness.

very notion of falsity—being ψευδής—carries overtones of deception and illusion.<sup>337</sup> This suggests that a promising way to unpack this notion is to say that a pleasure is false just in case it deceives us about some way the world really is.<sup>338</sup>

While in the grip of such false or deceptive pleasure, one is in the grip of a misrepresentation of the world: the representational elements that constitute one's pleasure misrepresent some way the world outside oneself really is. My deceptive pleasure might involve the false belief, say, that I am drinking a fancy wine, although it is in fact simple *vin de table*, or it might involve the false belief that money is a final good, although it is merely instrumentally valuable. In both cases, how the world strikes me and how the world really is are not aligned—appearance and reality come radically apart.

What this suggests, then, is that the subject of a false or deceptive pleasure is out of contact with reality—he lives in what Iris Murdoch calls 'a private dream world'—even though this tragic predicament cannot be appreciated from an inside perspective. By its nature as a representational state highly similar to belief, a pleasure necessarily presents its content as *true*: it suggests that some state of affairs obtains and that this state of affairs is good or otherwise positively evaluatively charged.

Dangerously enough, though, from an inside, first-person perspective, such a false pleasure cannot be told apart from a true pleasure that gets things right and informs us about the way things really *are*. Like all illusions, the pleasure-seeker's predicament—the fact that he is living in a private dream world—cannot be directly seen for what it is. From a subjective perspective, a true pleasure that gets things right and a false or deceptive pleasure that gets things wrong can after all not be told apart—in the same way in which a true belief

---

<sup>337</sup> See LSJ (s.v. 2).

<sup>338</sup> See especially Whiting (2014: 24) but also Fletcher (2018-b: 395–396).

and a false belief feel exactly identical from the inside.

This arguably explains why Socrates calls such a false or deceptive pleasure a ‘ridiculous imitation’ of true pleasure: if my pleasure lacks justification and if it does not make contact with some deeper reality beyond my head, even though I am not in a position to register this from my first-person perspective, I am less well off than I take myself to be—as Socrates’s later discussion of ridiculousness suggests.<sup>339</sup>

The proposal that false pleasures mislead us about some way the world really is makes trouble for the hedonist’s intuition that all and only pleasure matters or that pleasure is sufficient for a life worth living. Socrates’s jelly-fish *elenchus* had already forced Protarchus to concede that a life full of pleasure lacks other things of value—if only because ‘not possessing right belief, [we] would not realize that [we] are enjoying [ourselves] while we do (δόξαν μὴ κεκτημένον ἀληθῆ μὴ δοξάζειν χαίρειν χαίροντα).’<sup>340</sup>

By introducing the possibility of false pleasure and hedonic mistakes, I argue, Socrates introduces an analogous line of thought. If pleasure is in the business of representing the world, it seemingly follows that we do not just want to experience pleasure *tout court* but that we want our pleasures to latch onto the world and get things right. Given the fact that false pleasures are excluded from the final ranking, Plato seems to suggest, in other words, that pleasures only have value if they are true or reality-based.<sup>341</sup> Nozick describes the

---

<sup>339</sup> I am very much indebted here to Whiting’s (2014) excellent discussion of Plato’s defence of fallibilism in the *Philebus*.

<sup>340</sup> For discussion of this stretch of argumentation, see Evans (2007-b), Mouroutsou (2021), and O’Reilly (2019).

<sup>341</sup> If this conviction counts as a kind of puritanism—as Sumner (1996: 111) thinks it does—it suggests that even the puritan elements in Plato’s hedonic theorizing are philosophically motivated and rather sophisticated.

intuition his famous experience machine thought experiment is meant to trigger in a strikingly similar way: '[Another reason for not plugging into an experience machine is that] there is no actual contact with any deeper reality, though the experience of it can be simulated. Many persons desire to leave themselves open to such contact and to a plumbing of deeper significance.'<sup>342</sup>

## 5. Conclusion

In this chapter, I have examined Plato's Fallibilism Argument. In it, Socrates develops weaker and stronger types of hedonic cognitivism to defend hedonic fallibilism and resist Protarchus' claim that pleasure is immune to error. On a weaker line of thought, our pleasures are often predicated on, or accompanied by, a false belief which suggests that we cannot experience the pleasure at hand without being in error about some way the world really is. On a stronger line of thought, a pleasure is not just predicated on a (potentially false) belief, it is constituted by (potentially false) representational elements which suggests that the pleasure *itself* can be true or false.

In addition, I have argued that true and falsity is always a matter of (mis)representation: a pleasure is true if it gets things right and represents the world correctly, it is false if it gets things wrong and represents the world incorrectly. I have also argued that pleasure involves both a factual belief and an evaluative belief which suggests that we should recognize two ways in which pleasure can be false: a pleasure can be false if it claims that some state of affairs obtains,

---

<sup>342</sup> Nozick (1974: 43). Interestingly, the contemporary hedonist Fred Feldman (2004: 109–114) attempts to circumnavigate this problem by introducing the notion of 'truth-adjusted hedonism' on which a pleasure is only worth having if it is true and makes contact with the world.

although it does not, and a pleasure can be false if it claims that this state of affairs is valuable, although it is not.

Lastly, I have suggested that, taken together, hedonic cognitivism and hedonic fallibilism put pressure on Protarchus's hedonistic claim that pleasure is sufficient for the good life. For one, hedonic cognitivism serves as a neat illustration of the dialogue's claim that pleasure and cognition cannot be pulled apart: our pleasures derive their content from some cognitive operation of the human mind. For another, the fact that pleasures can get things right or wrong in virtue of their representational content triggers the intuition that we want our pleasures to make contact with a deeper reality which invites the conclusion, in turn, that human beings care about other things besides pleasure.<sup>343</sup>

---

<sup>343</sup> The material contained in this chapter has been presented in London (*British Society for the History of Philosophy Annual Conference*) and in Oxford (*Curing Through Questioning: Philosophy as Therapy Across Ancient Traditions and Modern Applications*) (special thanks go to Sybilla Pereira for amazing discussion of (Platonic) philosophy's therapeutic potential and psychotherapy's (hidden) philosophical underpinnings).