Descartes’ God is a Deceiver, and that’s OK

Abstract

That Descartes’ God is not a deceiver is amongst the canonical claims of early modern philosophy. The deep-cutting significance of this (purported) fact to the coherence of Descartes’ system is likewise canonical, infused in both how we teach and think about the project of the Meditations. Prevalent as it is, both ends of this narrative are suspect. We argue that Descartes’ color eliminativism, when coupled with his analysis of the cognitive structure of our sensory systems, entails that God is a deceiver. It’s doubtful that Descartes recognized this, given his frequent insistence that God is not a deceiver, and the role this plays in his system. But we argue that this is a concession that Descartes can grant, so long as we are careful about the kind of deception at play—a kind of deception Descartes does seem to recognize, albeit not without some ambiguity. On our story, Descartes’ metaphysics and epistemology are not driven by concerns about deception per se, but by concerns about God’s benevolence.

‘You say that God cannot lie or deceive. Yet there are some schoolmen who say he can’.
- Mersenne
1 Introduction

We’re often told that Descartes’ God is not a deceiver. We’re also told that God’s non-deceptive nature secures the metaphysics and epistemology that is foundational to Descartes’ philosophical system. That’s the prevailing wisdom. We aim to challenge it: in Descartes’ system, God is a deceiver. And that’s perfectly OK.

We argue that Descartes must endorse:

**Sensory Deception:** Many sensations are deceptive such that they are apt to produce false judgments, and their being so apt isn’t due to our violating various epistemic norms or the mere finitude of mind-body composites.

We can bridge Sensory Deception and God’s being a deceiver in three main steps. The first concerns the benign observation that bodies look colored. We have sensory experiences as of colored bodies. The second concerns Descartes’ denial that bodies are colored, namely, his commitment to color eliminativism.\footnote{While our argument focuses on color, it does not rely on color; the constraint is eliminativism about sensory qualities, not color eliminativism. The focus on color has theoretical and practical motivation: theoretical because color experiences deliver arguably the least ambiguous phenomenological predictions (making assessments of non-veridicality more tractable), and practical because Descartes had more to say about color than other sensory qualities and our experiences thereof.} Putting these first and second points together, it follows that color sensations are misleading. The third concerns a class of sensory judgments to the effect that bodies are colored that are immediately occasioned by color sensations (Hatfield 1986; Simmons 2003; Patterson 2016). These judgments are not just false; they are irrevocable, and our making them is automatic, and so not a matter of some shortcoming on our part, epistemic or otherwise. Since our sensory faculties are ultimately a product of God’s design, Sensory Deception prima facie implies that God is a deceiver.
There is a temptation to be dismissive here. One will insist that no matter how the details are filled in, this argument is flawed, or SENSORY DECEPTION is itself false. Never mind that Descartes often flatly denies that God is a deceiver. God cannot be a deceiver, the thought goes, because God is omniperfect. And with respect to our sensations, Descartes claims that God was benevolent in giving them the function of preserving our health (Sixth Meditation, AT VII: 83/CSM II: 57), and giving us the best of all possible sensations to that end (Sixth Meditation, AT VII: 87-8/CSM II: 60).²

Yet there is certainly no immediate tie between being omniperfect and being a non-deceiver, and nor is there an immediate tie between a type of sensation being optimally suited to preserve our health and its being non-deceptive. So these considerations fall far short of being decisive. On the other hand, if God were the cause of a malicious or harmful form of deception, that would be deeply problematic, so much so that this dismissive stance would be warranted.

This latter observation leads to two claims. First: if God is a deceiver, he better not be a malicious deceiver. Second: it’s question-begging to assume that any form of deception is incompatible with God’s omniperfection. Not everything that is deceptive is maliciously deceptive. And color sensations, we’ll claim, are such a case. More specifically, color sensations are non-intellectually, non-maliciously deceptive; non-intellectual because they are only partially grounded in the intellect (a body is needed too—see Principles I.32, AT VIII A: 17/CSM I: 204), and non-malicious because their being deceptive does not cause harm.³

Our precisified thesis, then, is that God is a non-malicious deceiver, and

²We employ the following abbreviations: ‘AT’: Oeuvres de Descartes (cited by volume and page), Adam and Tannery (1996); ‘CSM’: The Philosophical Writings of Descartes (cited by volume and page), Cottingham et al. (1985); ‘CSMK’: The Philosophical Writings of Descartes (cited by page), Cottingham et al. (1991).

³As we will see, there might be other ways in which deception conflicts with God’s omniperfection—e.g. via potential conflicts with God’s omnipotence. However, our focus will be on the relationship between malice, harm, and omnibenevolence.
that if Descartes were to admit this, it would be perfectly OK.\textsuperscript{4} Our case will begin with a brief overview of color eliminativism and its place in the secondary literature. Next, we argue that Descartes’ views on the nature of our sensory systems entails that God is a deceiver. We then look at four objections to this argument—three concerning whether we may in some way be responsible for our false color judgments, not God, and one concerning the unique kind of color judgments at play—arguing that these all fail. One might wonder how the very idea of non-malicious divine deception in sensory experience could get traction without at least assuming from the get-go that, for Descartes, an omnibenevolent God deceiving us is a live option. But this gets the dialectic backwards. The question is not how Descartes’ views about God may or may not constrain his views about our sensory systems. The question rather is what Descartes in fact says about our sensory systems; and what he says, on our view, entails that God is a deceiver.

In view of this, one might just say that, even if our argument is sound, Descartes was simply inconsistent, not realizing what his views about our sensory system committed him to. That’s certainly plausible, and not an option we attempt to rule out here. Instead, what we attempt to do is explore whether Descartes could have allowed for divine deception, irrespective of whether he actually did, and what this might look like. Without making any claims about what Descartes definitively meant, we’ll suggest that Descartes has the resources to allow for our ‘omni-compatible’ sense of divine deception, while maintaining the integrity of his core arguments in the Meditations. As Mersenne tells Descartes, various medieval philosophers allowed for divine deception (Perler 2010); we suggest that Descartes can too.

\textsuperscript{4}Sarah Patterson (2016) has also recognized that SENSORY DECEPTION is true through a general analysis of sensory experience (although she does not focus on color), and that SENSORY DECEPTION poses a problem for God’s non-deceptive nature. However, she claims that Descartes does not have any viable solution to this problem. We do: Descartes can admit that God is a (non-malicious) deceiver.
2 Color Eliminativism and Deception

2.1 Color Eliminativism

Bodies—tables, grass, and national flags—look colored. We also think that, absent the possibility of an illusion, bodies have the colors they look to have.

The color eliminativist—or simply ‘eliminativist’—says that this is wrong. Bodies look colored. We have color sensations. But bodies are not colored. Eliminativism is thus an error theory: color sensations misrepresent external bodies as colored.\textsuperscript{5} And this is systematic. Color sensations are not just occasionally non-veridical; they are systematically non-veridical by their nature.\textsuperscript{6}

There are different versions of eliminativism—differing, for instance, on whether colors are uninstantiated, or just not instantiated where they appear to be—but these details won’t matter. What matters is that Descartes was an eliminativist of some stripe. And we will assume as much. This assumption is well-founded. Eliminativism is part and parcel of the early modern philosophical ethos. Galileo, Descartes, Hobbes, Boyle, Malebranche, and Locke were mechanists about bodies, making eliminativism all but a foregone conclusion (cf. Chamberlain 2019a).\textsuperscript{7}

\textsuperscript{5}We’ll assume representationalism (Simmons 1999, De Rosa, 2007; Hatfield 2013), but qualia realists (MacKenzie 1990) can grant eliminativism too.

\textsuperscript{6}The eliminativist is concerned with our ordinary recognitional concept of color (Maund 2011)—sometimes called ‘colors-as-we-see-them’ (Mackie 1978), ‘phenomenal colors’ or ‘sensuous color’. Chamberlain (2019a: 294) uses ‘sensuous color’, defined as the property that “material things visually appear to have when they appear to be colored.”

\textsuperscript{7}See Principles II.4 ((AT VIII A: 42/CSM I: 224). For a sampling of eliminativist readings of Descartes see: Buroker (1991), Chamberlain (2019a), Cook (1996), Downing (2011), Garber (1992: 75), Keating (2004), Maund (2011: 382), Rozemond (1999: 452-3). Indeed, even readings that, on their surface, appear to cut against an eliminativist view, are in fact eliminativist. For example, Cottingham (1989) argues that, according to Descartes, bodies have dispositional colors. But this does not make Descartes a dispositionalist, because Cottingham (Ibid.: 238) denies that dispositional colors are \textit{colors}; the eliminativist will concur with this stance, but that’s not to say that she must deny that bodies have dispositions to experientially affect certain perceivers in certain contexts. Likewise, Ather ton (2004) argues that, according to Descartes, bodies have physical colors, but she also denies (Ibid.: 33) that physical colors are \textit{colors}; the eliminativist will concur with this
Now Nolan (2011) argues that Descartes is a color nominalist. Color terms don’t express properties of minds or bodies. This is compatible with eliminativism, since the eliminativist can deny that colors exist at all, even as uninstantiated properties (Hardin 1988). Nolan also insists that color appearances are not sensory or perceptual, but wholly due to habitual (false) judgments. This too is compatible with eliminativism, but as it’s a point that will be central in our argument to follow, it’s worth flagging now. Nolan’s motivations here are multi-faceted, but of note in the present context is his (2011: 98) insistence that letting anything but habitual judgments attribute colors to bodies will tell against God’s non-deceptive nature. For if the misleading color appearances are in part due to false judgments—habitual judgments which are revisable—then God is off the hook.

We’ll return to this issue in Section 2.2.2, but one point will suffice for now. If these judgments are revisable, we can ask what happens when we revise them. And once we ask that question, we get a plainly false prediction: once the judgments are revised, bodies should no longer look colored. But that’s not what happens. I can stop judging that a stop sign is red—perhaps because I now believe that eliminativism is true—but this will have no bearing on how the stop sign looks to me. It’s not as if once one becomes an eliminativist, one no longer agrees with non-eliminativists about the appearances of things. Surely Descartes would have recognized this.

So Nolan should allow that color appearances are sensory. This does not mean that color appearances are not always accompanied by judgments; indeed, on our analysis of Descartes, they are. But these judgments are not revisable. Nolan is right to worry about the implications of this claim for God’s non-deceptive nature. And of course that’s exactly our point.

stance, but that’s not to say that she must deny that bodies token spectral reflectance types (cf. Byrne & Hilbert 2003).
2.2 God is a Deceiver

Descartes only offers examples of deception (see, e.g., First Meditation, AT VII: 18/CSM II: 12, Discourse I, AT VI: 9/CSM I: 115, Optics AT VI: 145/CSM I: 174). Even so, given his examples, Descartes would endorse:

**Deception:** $X$ deceives a subject, $S$, if (i) $S$ judges that $p$; (ii) it’s false that $p$; (iii) $X$ judges that $p$ is false; (iv) the evidence $E$ on the basis of which $S$ judges that $p$ is brought about by $X$ for the purpose of making $S$ judge that $p$.

Deception captures a form of deception by commission.\(^8\)\(^9\) However, as we will see, whether God is a deceiver under Deception will depend on how ‘judge’ (and ‘judgment’) is understood. Thus Deception needs to be unpacked thoroughly.

Our argument runs as follows:

D1 Bodies are not colored.
D2 In virtue of bodies looking colored—i.e. in virtue of the fact that certain phenomenological evidence $E$ obtains—the subject, $S$ judges that bodies are colored.
D3 So, in virtue of bodies looking colored—i.e. in virtue of the fact that certain phenomenological evidence $E$ obtains—$S$ falsely judges that bodies are colored. [D1-D2]
D4 God brought about $E$ for the purpose of making $S$ falsely judge that bodies are colored.
D5 God knows that bodies are not colored.
D6 If D3, D4, and D5, God is a deceiver [given Deception].

---


\(^9\)Dropping ‘for the purpose of’ from clause (iv) makes room for non-intentional forms of deception (Chisholm & Feehan 1977). Everything to follow will apply on non-intentional deception too, since (iv) entails its weaker non-intentional, purely causal cousin.
God is a deceiver.

D1 follows from eliminativism. D5 is an assumption. Since D3 is a bridge conclusion, what remains to be discussed are D2, D4, and D6.

2.2.1 Regarding D2

D2 concerns sensory judgments. Broadly writ, a ‘sensory judgment’ refers to a judgment about some sensory state of affairs. A sensory state of affairs can concern just sensible qualities (for example, the redness of a rose), sensations themselves, or both. Not all sensory judgments need to be immediately caused by a sensation. I could, for instance, judge that roses are red based on my memory of seeing a rose. However, in the Sixth Replies Descartes recognizes a class of sensory judgments—one of three grades of ‘sensory response’—that by their nature are immediately caused by sensations. The first grade of sensory response is purely physiological and consists of the “immediate stimulation of the bodily organs by external objects” and terminates in a brain impression (AT VII: 436/CSM II: 294). The second-grade of sensory response occurs in the mind, and comprises “all the immediate effects produced in the mind as a result of its being united with a bodily organ which is affected in this way” (Ibid.). These are ‘sensations’ in our parlance. The third-grade of sensory response involves cognitive processes, and “includes all the judgments [judicia] about things outside us [de rebus extra nos] which we have been accustomed [consuevimus] to make from our earliest years—judgements which are occasioned by the movements of these bodily organs” (AT VII: 437/CSM II: 294-5).

The operative word here is ‘occasioned’. Let third-grade judgments be judgments about sensory states of affairs—specifically, sensible qualities—that are immediately occasioned by events at the first and second-grade of sensory response. We think third-grade judgments have two unique features. First, they in some way structure how things look phenomenologically. Second, they are irrevisable. Following Simmons (2003), we will refer to third-
grade judgments as *constructive judgments* (‘c-judgments’) to emphasize this first feature. So defined, not all sensory judgments are c-judgments. Some sensory judgments might neither structure phenomenology nor be irrevisable. These are *reflective judgments* (‘r-judgments’). The existence of r-judgments should be uncontroversial, but c-judgments also have gained broad favor in the secondary literature. Simmons (Ibid.), Hatfield (1986), and Patterson (2016) all contend that Descartes countenanced c-judgments.¹⁰

Now D2 says that $S$ judges that bodies are colored in virtue of bodies looking colored (i.e. E). While we often do make r-judgments to this effect, the sense of judgment relevant to D2, and our case that God is a deceiver, concerns c-judgments. D3 (that $S$ falsely c-judges that bodies are colored) follows from D1 and D2. We will thus take up three claims: first, that Descartes countenanced c-judgments; second, that c-judgments about color are false; and third, following Patterson (*pace* Simmons) c-judgments are issued by the will. A corollary of the third claim is that c-judgments, if not beliefs, are very much belief-like.

**Claim One.** Let’s start by looking again at Descartes’ three grades of sensory response, focusing on color. At the first grade, there is some property in bodies—perhaps a dispositional property or the categorical basis of this dispositional property—that leads to a brain impression that, in turn, causes a color sensation. The sensation itself is the second-grade of sensory response. It’s in virtue of this second-grade response that bodies *look* colored. This is because sensations *represent* bodies as colored. That sensations represent follows from Descartes’ general claim that all mental states have objective reality, that they are “of things [rerum]” (Third Meditation, AT VII: 44/CSM II: 30), and his more specific claim that sensations inform [*significandum*] (Sixth Meditation, AT VII: 83/CSM II: 57). Given eliminativism, this implies that bodies are misrepresented as colored. So, again, when it comes to D2, and its talk of ‘phenomenological evidence’, this concerns the

¹⁰When unqualified, ‘judgment’ can refer to a c-judgment or r-judgment.
second-grade of sensory response; it’s the misrepresentation at the second-grade that constitutes the phenomenological evidence (E) in virtue of which S is deceived, if in fact S is deceived.

At the third-grade, the mind makes a host of judgments about what occurs at the second-grade. The question is whether these judgments structure sensory phenomenology and are irrevisable—that is, whether there are c-judgments so defined. Here, two points are especially germane:

(1) Third-grade judgments partially structure sensory phenomenology.

(2) If third-grade judgments partially structure sensory phenomenology, then given Descartes’ views about the status and bio-functional role of our God-given sensory system, we should expect that third-grade judgments are also irrevisable.

If (1) and (2) hold, then third-grade judgments are c-judgments so defined.

Extensive arguments for (1) have been offered by Simmons (2003), Hatfield (1986), and Patterson (2016), so we’ll confine ourselves to two clarificatory remarks. First, most of the examples of c-judgments adduced by Simmons, Hatfield, and Patterson pertain to size, shape, and distance. However, considerations of theoretical economy suggest that if Descartes countenanced judgments that partially determine sensory phenomenology at all, then he countenanced them for all properties represented at the second-grade, and so colors too. Indeed, on the assumption that Descartes is talking about

---

11 As just one example, in the Optics, Descartes first explores the idea that judgments are involved in sensory phenomenology:

When we look at two ships out at sea, one smaller than the other but proportionately nearer so that they appear equal in size, we can use the difference in their shapes and colours, and in the light they send to us, to judge which is the more distance. (AT VI: 140/CSM I: 172)

A third-grade judgment about the distance between the ships—based on the shapes and colors represented at the second-grade—in part determines the appearance of the ships’ distance. According to Simmons, “[t]hese judgments associate such things as brightness with proximity, and dimness with distance, so that the brighter ship looks to be closer in the resulting sensory experience” (2003: 572).
such judgments in the aforementioned passage of the *Sixth Replies*, he makes this point rather clearly, saying that these judgments include “all judgments about things outside us.” And so, since color sensations concern things outside us (bodies), there will be judgments that structure the phenomenology of color experience if there are such sensory judgments at all.

Second, we make no claims about how exactly these judgments structure phenomenology. In the case of judgments about shape, Simmons tells us that:

> When I look at a bagel from an oblique angle, I judge that it is round (or perhaps toric) despite the elliptical patch of brown it produces in my visual field at the second-grade of sensory perception; the result is a sensory experience in which the bagel looks round (or toric). (2003: 554)

With color sensations, Simmons suggests that these judgments account for color constancy amongst bodies, representing that particular bodies instantiate a particular color across distinct illumination conditions (2003: 573). As Simmons writes: “I may therefore judge, at the third-grade of sensory perception, that those leaves are green with the result that the leaves look like green leaves in changing light” (Ibid.; emphasis added). We don’t know if Simmons is right about this. Our claim is just that there are sensory judgments that structure the phenomenology of sensory experience, and this includes sensory experiences of color.

Regarding (2), to say that a third-grade judgment is irrevisable is to say that one cannot cease the formation of a third-grade judgment, or alter its content-type, once a second-grade sensation is had. To say that they are revisable is to say otherwise. The case for (2) is broadly theoretical. The case, inspired by Patterson (2016), goes like this. Descartes claims that our sensory faculties—including all grades of sensory response—have a biological function of preserving the health of mind-body composite. As such, it’s not just the sensations at the second-grade that play this role, but third-grade judgments as well. And critically, while both the second and third-grades
work in concert to preserve our health, the \textit{way} they do this will be different. \textit{Whatever} phenomenology-determining role third-grade judgments play—and we know that they play some role, given point (1)—it’s \textit{that very role} that helps preserve our health. For example, if Simmons is right, the second-grade makes bodies look colored \textit{simpliciter} (something that is good for us), and the third-grade makes bodies look to have the \textit{same} color across different illumination conditions (also good for us). More generally, if God gave us the best of all possible sensory systems, then the way third-grade judgments structure phenomenology must be perfectly suited towards our preservation. In other words, the third-grade judgments \textit{we in fact} issue have to be the best of all possible such judgments, \textit{qua} their bio-functional role. If God gave us the best of all possible sensory systems, and such judgments are part of that system and so figure into the preservation of our well-being, then any attempt to revise them as fallible, finite beings would \textit{ipso facto} make them worse. Thus from a design standpoint, it makes more sense for these judgments to be irrevisable. It’s no wonder then that they are considered \textit{automatic}, and caused directly by sensory states without any inference. Anything short of this would, we presume, permit them to be revisable.\textsuperscript{12}

So Descartes plausibly countenanced c-judgments—i.e. irrevisable sensory judgments that partially structure phenomenology. Of course, if I wear color-tinting glasses I ‘revise’ my sensations and c-judgments, but this is only at a \textit{token} level. With color-tinting glasses, my color sensations are of the same \textit{type}; they still represent bodies as colored. Likewise for my c-judgments; they

\textsuperscript{12}Simmons take on irrevisability is different. For Simmons, c-judgments are an essential constituent of sensory experience, and since sensory experiences are themselves irrevisable, she concludes that third-grade judgments are irrevisable (2003: 578). We are not inclined towards this line, since we think there is a good sense in which the second-grade constitutes a sensory experience all on its own. Simmons reserves ‘sensory experience’ for the joint product of the second-grade sensation and a c-judgment. We reject this usage. This is not to deny that c-judgments have a unique role, as Simmons contends. Since on our view the second-grade sensation is sufficient for a body to look \textit{colored} \textit{simpliciter}, this strikes us as good of a candidate as any for being a sensory experience. What we insist on is that there is no \textit{further} thing that results when second-grade sensations and c-judgments co-occur.
are judgments which (in some way) represent bodies as colored. All color-tinting glasses would do is change which color property is represented. At the type level, revision is impossible. For if our sensory system really is the best, any revision at that level would ipso facto make them worse.

R-judgments, by contrast, are revisable (Simmons 2004: 567-8; cf. Patterson 2016: 99). They are formed through inference. The meditator judges that “the heat in a body is something exactly resembling the idea of heat which is in me” (AT VII: 82/CSM II: 56). This judgment is revised—I can cease to have it altogether—once the meditator learns that heat isn’t a property of extension. R-judgments don’t structure sensory phenomenology, and so are not part of our sensory system in the strict sense.\[13\]

**Claim Two.** Descartes and commentators are clear that all c-judgments are false (Sixth Replies, AT VII: 437-439/CSM II: 295-6, Simmons 2003: 573-4, Patterson 2016: 101). If that’s right, in the case of c-judgments about color, their content can’t be that bodies merely look colored. They will at the very least include the claim that bodies are colored. This claim is false, given eliminativism. And since c-judgments are *occasioned* by color sensations, we should expect robust similarity in content between the latter and the former. So it’s no surprise that c-judgments have the false contents that they do.

**Claim Three.** For Simmons, r-judgments are not just different from c-judgments because the former, but not the latter, are irrevisable. They are also different because the latter are not beliefs, even if they often lead to them (Simmons 2003: 567). A judgment is a belief only if it is issued by the will. Since Simmons denies that c-judgments are issued by the will, she claims they are not beliefs. Rather, they are a product of the intellect.

As to the *source* of c-judgments, Simmons is wrong. She realizes that

\[13\] This point of comparison between c-judgments and r-judgments roughly tracks the distinction between ‘perceptual beliefs’ and ‘central beliefs’ familiar from contemporary discussions of perceptual-belief formation and dual-component views of sensory experience (e.g. Quilty-Dunn 2015). On the similarly “coercive”, “cognitively spontaneous” nature of perceptual beliefs, see BonJour (1985: 117).
her interpretation conflicts with Descartes’ standard theory of judgment in the Fourth Meditation, according to which judgments are issued by the will (2003: 565-6). Simmons supports this alternative reading by drawing from Descartes’ claim in the *Sixth Replies* that “although such reasoning is commonly assigned to the senses (which is why I have here referred it to the third-grade of sensory response), it’s clear that it depends solely on the intellect [*a solo intellectu pendere*]” (*Sixth Replies*, AT VII: 437-8/CSM II: 295; cf. *Optics* (AT VI: 137-8/CSM I: 170).

However, as Patterson rightly points out, Descartes’ use of ‘intellect’ in this context is misleading:

I read Descartes’s talk of the intellect alone as designed to emphasise that the senses are not involved, rather than to exclude any role for the will. This reading gains support from the fact that Descartes associates judgments in the full-blooded sense with the ‘intellect alone’ at the end of the Second Meditation, where his point is also to contrast judgment with the senses and imagination. . . Here Descartes uses ‘intellect’ as an umbrella term to cover intellect and will, the faculties of pure mind, when a contrast is being made with the faculties of the embodied mind. (2016: 85, fn. 48)

Thus we’ll assume r-judgments and c-judgments are issued by the will.14

---

14This assumption might engender two worries, one concerning Descartes’ activity versus passivity distinction, and one concerning freedom. On the former, Descartes’ claim that sensations are wholly passive is consistent with c-judgments being issued by the will. This *prima facie* inconsistency is in part due to Descartes’ slippery usage of ‘sensation’. As evidenced by his clarifications in the *Sixth Replies*, he sometimes uses ‘sensation’ to refer to the composite sensory experience that includes the third and second-grade of sensory response, and sometimes he uses ‘sensation’ to refer to the second-grade alone (our usage). Resolving the tension is straightforward: sensations *at the second-grade* are wholly passive. On the latter, the worry is that if we cannot help but issue c-judgments, then the will cannot be wholly free. While this matter is tricky, it’s fairly clear that Descartes recognizes *conditional* constraints on the will. For instance, Ragland (2006) contends that Descartes endorsed this claim: given certain conditions, if *S* has a clear and distinct perception that
On the relationship between c-judgments and beliefs, if being issued by
the will were sufficient for being a belief, then pace Simmons, c-judgments
would be beliefs. It will not be useful to adjudicate here what really counts
as a belief. So we will be agnostic on this sufficient condition. That said, we
know that beliefs have propositional content, are assertoric in force, capable
of figuring in inferences, linguistically expressible in suitable creatures, and
possessive of a distinctive functional profile—that is, they interact in char-
acteristic ways with desires, emotions, action, and memory (Quilty-Dunn
2015: 559). C-judgments don’t obviously tick all of these boxes. For exam-
ple, Simmons (2003: 554 fn. 13) flirts with the idea that c-judgments are
at least confusedly presented to consciousness (Principles I.46, AT VIII-A
22), which might make them linguistically inexpressible. Yet as Quilty-Dunn
notes (2015: 559), given that beliefs are functional states, and so specified in
terms of roles they are disposed to fulfill, that c-judgments may not actually
fulfill these roles is otiose: “dispositions can be muted” (Ibid.).

The status of c-judgments as beliefs, then, is unsettled. And we are happy
to keep it at that. What we do insist on, however, is that c-judgments are
certainly ‘belief-like’; they are propositional and they have assertoric force,
and they are truth-apt. Following Cassam (2010: 83), judging that p is “the
act of occurrently putting p forward in one’s mind as true” (2010: 83). So
one can thus c-judge falsely, i.e. put p forward in one’s mind as true when
it isn’t. And (falsely) c-judging that p, if not requiring one to believe that p,
will at least dispose one to believe that p.

Going forward, then, the primary takeaways are two-fold. First: when it
comes to D3, and its talk of ‘false judgments’, this concerns c-judgments
not r-judgments. Second: c-judgments are issued by the will, are at least

\[ F, \ S \text{ cannot refrain from judging that } F \ (\text{Fourth Meditation, AT VII: 58-9/CSM II: 41).}
\]

Yet this is structurally analogous to our claim about c-judgments: we only claim that if
\[ S \text{ has a second-grade sensation, then } S \text{ cannot refrain from issuing a c-judgment. And}
\]

as before, the reason has to do with optimal design: God deemed this set-up best for the
preservation of our health.
belief-like, and are automatic and irrevisable. And as will become clear, the existence of c-judgments is what renders Descartes’ God a deceiver. What is deceptive are second-grade color sensations, but $S$ is only deceived once she makes false c-judgments.

2.2.2 Regarding D4

D2 claimed that $S$ makes c-judgments to the effect that bodies are colored due to certain phenomenological evidence $E$—i.e. the phenomenological fact that bodies look colored at the second-grade. D4 goes further. Adjusted for subsequent clarifications, it says:

\[ D4 \quad \text{God brought about } E \text{ for the purpose of making } S \text{ falsely c-judge that bodies are colored.} \]

D4 corresponds to clause (iv) of Deception. By ‘brought about’ we mean that God is broadly causally responsible for the makeup of our sensory system, which includes the way things look. We see three ways in which D4 might be false. Each objection targets clause (iv) by denying that God is responsible for our deception, but they do so differently. The first objection trades on our finitude; God designed us and our sensory systems, but without any intent _per se_ to deceive us; bodies look colored because of a design glitch. On the second objection, _we ourselves_ generate the misleading evidence via our habitual violations of epistemic norms. On the third objection, given our violation of epistemic norms, while $S$ does not generate the evidence, $S$ misunderstands what the second-grade is telling her. And it’s only because of this misunderstanding that $S$ is deceived; once the norm is followed, $S$ realizes that the second-grade isn’t misleading after all. The common thread of these objections is that although $S$ can be deceived, it’s not due to God. Responding to these objections substantiates Sensory Deception.

**Objection One.** The first objection to D4 leverages the Sixth Meditation response to the problem of dropsy (AT VII: 87-9/CSM II: 60-1). The thirst
sensations $S$ has when suffering from dropsy are both deceptive and harmful. Descartes responds by noting that God’s intention in designing our system was to give us the best of all possible sensations (Sixth Meditation, AT VII: 87-8/CSM II: 60). However, this optimal design is implemented in finite mind-body composites. And finitude allows for glitches in our sensory system. An aberrant motion in the body can cause a sensation that is harmful to us. Such glitches are due to our finitude, not God’s design of our sensory system (Sixth Meditation AT VII: 88-9/CSM II: 61). By extension, the thought then is that color sensations misrepresent bodies as colored due to our finitude.

This objection fails. Unlike dropsy, color sensations are systematically misleading. Dropsy is a rare or one-off affair, explainable by our finitude and the failure of proper correspondence between physiological stimulation, brain impression, and sensation. By contrast, the error inherent to color sensations leaves every impression of design; bananas standardly look yellow (but are not), leaves standardly look green (but are not), and they all do so in predictable circumstances.

This point is worth embellishing. In general, if I want to know what an agent’s intentions were in creating some system, then I should examine how it behaves under optimal conditions. Suppose you had a typewriter where, every time you hit the ‘H’ key, a ‘P’ was printed on the page. The typewriter is clearly not working correctly. Because of this, you shouldn’t infer at the outset that the designer of the typewriter wanted you to spell ‘horses’ as ‘porses’ and ‘hats’ as ‘pats.’ There has been a mistake. Similarly, the fact that dropsy rarely occurs, and only occurs when the body isn’t functioning properly, tells against God’s bringing about dropsy so that $S$ would drink. We’ll return to this point later (Section 3.2), but the point for now is that we have precisely the opposite situation in the case at hand: the systematic illusoriness of color sensations is not a glitch. Things are working precisely how they are supposed to be working. Our visual system is supposed to falsely attribute color properties to bodies. It’s reasonable to infer, then, that it was
a part of God’s design plan.

**Objection Two.** The second objection leverages the Fourth Meditation account for how to avoid intellectual error:

**Truth Rule:** A subject, $S$, should not judge that $p$ unless $S$ is currently clearly and distinctly perceiving that $p$. (Third Meditation, AT VII: 35/CSM II: 24; Fourth Meditation, AT VII: 60-1/CSM II: 41)

Since color sensations are not clear and distinct, when $S$ affirms them she violates the Truth Rule. I improperly use my will by extending it to that which is obscure and confused. Therefore, the source of the error lies in us, not God. So, God is not a deceiver.

The Truth Rule, as described in the Meditations, is restricted to r-judgments. For the Truth Rule to have any teeth, it must apply to judgments that we can revise or cease to have. So it cannot apply to c-judgments. But it can still be apt, and relevant to D4, if appropriately supplemented. The supplementation draws a connection between judgments we can revise—i.e. r-judgments—and how bodies look at the second-grade. This is critical, because even if we do violate the Truth Rule, we violate it because of the misleading phenomenological evidence at the second-grade—evidence, that has, on the face of it, nothing to do with us. Here is the combined principle:

**Fault:** Bodies look colored only when $S$ fails to exercise her will properly—that is, when $S$ r-judges in a way that violates the Truth Rule.

Fault says that when $S$ judges *in a way where she could have judged otherwise*, she creates the misleading appearance of bodies looking colored. Indeed, as we saw, Nolan (2011: 98) endorses something like Fault in his attempt to acquit God of deception. The idea is that the misleading phenomenological evidence at the second-grade is somehow due to $S$ *in the first place*.

This move is a non-starter. First, recall that there are principled reasons—viz. that God gave us the best sensory system—that the second-grade is itself
stable and irrevisable. This objection is predicated on such an assumption being false, because it turns on $S$ being in some sense able to correct the type of sensations she has (e.g. sensations which predicate colors of bodies).

But suppose we suppress that assumption. The objection still fails. For when we consider the relevant sense of ‘looks’, Fault is false; bodies still look colored even when I correctly apply the Truth Rule. Compare:

(a) The book looks (seems) red to $S$.

(b) It looks (seems) to $S$ like the kitchen was ransacked.

‘Looks’-statements (or ‘seems’-statements) are utterances of sentences like (a) and (b). But the sense of looks in (a) is quite different from that in (b). Sentence (a) employs a non-comparative phenomenal sense of ‘looks’, whereas (b) employs an epistemic sense of ‘looks’ (Chisholm 1957; Jackson 1977). When ‘look’ is used epistemically, the sentence conveys what is subjectively probable, conditional on the evidence available. A mark of epistemic looks is that they go away in the presence of a defeater if the agent is rational. So, if you are rational, it will stop looking as if a hungry burglar ransacked your kitchen once you see the paw prints and deduce that it was a raccoon.

That’s not so with when ‘look’ is used phenomenally. Recall (Section 2.1) the discussion of Nolan’s nominalism. Even when $S$ learns that she is suffering an illusion, and that the book is really green, she will still sincerely utter (a). This point is not limited to color. Consider the Müller-Lyer illusion:

![Fig. 1: The Müller-Lyer Illusion](image)

The second line does not stop looking longer than the first even when I know that they are equal in length. But now notice: relevant for Fault is the
phenomenal sense of ‘looks’—the sense in which bodies look phenomenally colored. Thus, Fault is false, and in a way that anyone can see, Descartes included. This is not like dualism.\footnote{Müller-Lyer type cases—or known-illusions—are compatible with the present story about c-judgments. When the subject denies that the book is red post-Meditations, she is expressing her lack of endorsement for her c-judgment that says otherwise. This is not surprising given the non-inferential nature of c-judgments, and our more general recognition of having cognitive states that we fail to endorse (Quilty-Dunn 2015). Issues related to this point will become central in Section 2.2.3.}

**Objection Three.** The third objection to D4 does not claim that $S$ generates $E$. Instead, it claims that, given $S$’s violation of another epistemic norm, $S$ misunderstands $E$; $S$ is wrong about the ‘claims’ sensations make about the world. As such, insofar as $S$ is deceived by sensations, the mistake is on $S$, not God.

Simmons (1999) advocates something like this move. Whether a sensation will give rise to a true or a false r-judgment depends on what $S$ believes about the function of the sensation. Given their biological function, the idea is that sensations are not designed to guide us in metaphysical inquiry. As Descartes says, “I misuse them [sensations] by treating them as reliable touchstones [regulis certis] for immediate judgments [immediate dignoscendum] about the essential nature of the bodies located outside us” (Sixth Meditation, AT VII: 83/CSM II: 57-8). Or as Simmons puts it: “[w]hat is important here is that the level at which sensations get their representational hook on the world is the level of ecology not physics” (1999: 356). If $S$ understands the proper function of sensations, then sensations will guide $S$ to making true judgments about bodies in her environment. This is the sense in which sensations are materially true for Simmons (1999: 363). However, if we misunderstand the function of sensations—believing that they are accurate guides to the metaphysics of res extensa—then we will make false r-judgments about bodies. This is the sense in which sensations are materially false for Simmons (1999: 363; cf. Nelson 1997). In short, Simmons is providing another epistemic norm that is supposed to guide our r-judgments:
**Function Rule:** $S$ ought to clearly and distinctly perceive the essential biological function of a given sensation before making any r-judgments based on that sensation.

Function Rule, as a species of the Truth Rule, is only intelligible if it's restricted to r-judgments. But presumably the idea is that when $S$ r-judges without heeding the Function Rule, $S$ is apt to misinterpret what the second-grade is telling her. If $S$ were to follow the Function Rule she would see that her sensations are actually materially true. Yes, perhaps bodies don’t have colors. But our sensations don’t exactly say that bodies have them either—at least not in any deep metaphysical sense. As such, since $S$’s sensations, properly understood, are not deceptive to begin with, God cannot be a deceiver in virtue of giving us those sensations.\(^\text{16}\)

Two points. First, there is something obscure about the distinction Simmons is tacitly drawing. Simmons claims that sensations have a “representational hook” on the world (1999: 356). For a sensation to represent, it must make a ‘claim’ about the world. For Simmons, these claims are not metaphysical but “ecological”. It’s not exactly clear what justifies this distinction. Of course, we can recognize assertoric claims about the world that fail to have metaphysical import. For example, we might imagine a person who, upon looking at a bare cupboard, remarks that there is some empty space to place a box. That claim does not in itself commit the speaker to any metaphysical claims, since perhaps she does not include empty space in her ontology. Yet this analogy distorts how color sensations work, and indeed how they must work, given Descartes’ commitments about their function. Color sensations plausibly have an attributive structure, and they have this structure because it is by attributing color properties to bodies that we can navigate the world. We can think of their content as being of the form $<x \text{ is } F>$, where ‘$x$’ picks out an external body and ‘$F$’ picks out some color property. To say that noth-

\(^{16}\)Simmons (1999) employed this move before her introduction of c-judgments (2003). Our guess is that she would now reject D6. We address this objection next in Section 2.2.3
ing is colored is to say that there is no x that falls under the extension of F. But absent the quite rare cases where S is undergoing a total hallucination, there always will be some x.

In this way, color sensations on eliminativism are best thought of as systemic illusions, and it is a conceptual truth that, in illusions, we are aware of an existent object that is at variance with how it looks to be. This point is absolutely crucial, for in the empty space case, ex hypothesis, there is no object to which we are attributing a property. Given the person’s ontological commitments, there is nothing. Yet that’s not so in the color case. And that is really all that matters for color sensation’s making metaphysical claims. In attributing a property to an object, we are considering what that object is like. That, in some good sense of the term, is a metaphysical claim. This may not be a claim about what the object is like essentially, but not all metaphysical claims are claims about essential properties, and the same goes for Descartes’ system (they can be about modes). And the property may also only be relational—a property bodies have “relative to us and to our continued well-being” (Ibid., p. 356). Yet this too is metaphysical, since saying what bodies are like relative to us is still saying what bodies are like.

So by the mere fact that sensations represent bodies, they will in doing so make some sort of metaphysical claim about bodies. The deeper, second point following from this, though, is that when we consider how color sensations preserve our health, we see that these metaphysical claims are false. Whether or not bodies are colored, we have sensations as of colored bodies. As with all sensations, color sensations have the bio-functional role of preserving the health of the mind-body composite. And there can be no doubt that essential to realizing this function is their representing bodies so that they look colored. Consider that a primary function of color vision is the recognition of food sources, like ripe fruit. It’s at best difficult to see how S could recognize a ripe banana without representing it as yellow—a claim about what bananas are like. Once this point is recognized, however, and we consider it
in light of Descartes’ eliminativism, Simmons’ position becomes untenable. For notice: if an essential feature of the biological function of color sensations is the representation of bodies as colored, then any r-judgment made about bodies in our environment when considering color sensations *qua* their essential bio-functional role will at least include the claim *that bodies are colored*. However, if color sensations were materially true, then, presuming the Function Rule, any such r-judgment will be true. Yet given eliminativism, the claim *that bodies are colored* will *not* be true. Therefore, no r-judgment made about bodies when considering color sensations *qua* their essential bio-functional role will be true. In other words, color sensations are materially false. So *S* is *not* mistaken about E. D4 is safe.

Perhaps one might challenge the idea that these r-judgments will include the claim *that bodies are colored*. Alas, doing so in an ad-hoc manner is likely hopeless. Consider the r-judgment *yellow bananas are ripe*. That’s true. But aside from the fact that it’s not clear how to connect such an r-judgment in the appropriate way with our color sensations—after all, it’s hardly clear that sensations themselves attribute ripeness to bananas—a response here cannot not simply say that, in accordance with the Function Rule, *S* makes true r-judgments to the effect that bananas are ripe. It would also have to say that, in accordance with the Function Rule, *S does not* and *cannot* make false r-judgments to the effect that bananas are yellow. The problem though is that if the representation of bodies as colored is essential to the function of color sensations, the would-be proponent of this move would simply be *stipulating* that r-judgments to the effect that bananas are yellow are barred. The Function Rule considered on its own terms provides us with no underlying reason for why this should be so.

---

17 Whether sensations (strictly at the second-grade) represent the ripe banana as ripe is highly contentious, since it’s doubtful that our visual system proper is even in the business of representing anything beyond ‘low-level’ properties (size, shape, color, distance). See Hawley & Macpherson (2011).
2.2.3 Regarding D6

Recall D6: if D3, D4, and D5, God is a deceiver. It might seem like D6 isn’t in need of any defense. D5 states that God knows that bodies are not colored. D3 states that $S$ falsely c-judges that bodies are colored in virtue of certain phenomenological evidence E, viz. that bodies look colored. And D4, as we have just examined, says that God brought about E for the purpose of making us falsely c-judge that bodies are colored. As such, D6 follows given the definition of Deception. This is the line Patterson (2016: 98) takes, emphasizing as she does that if God gave us a natural propensity to make false c-judgments, then God’s being a deceiver is a genuine problem. Yet this inference can be blocked by arguing that Deception is only a sufficient condition for deception if its use of ‘judgment’ concerns r-judgments. And since r-judgments are revisable and governed by epistemic norms like the Truth and Function Rules—rules $S$ flouts—God is off the hook. If all goes well post-Meditations, then $S$ no longer r-judges that bodies are colored.

The objection can be fleshed out by considering again how the Truth and Function Rule were employed in the discussion of D4. There, the former was used—supplemented by Fault—to examine how E might be self-manufactured, while the latter was used to show how we might somehow be mistaken about E. But suppose instead we grant that God did bring about E and that E is deceptive—i.e. suppose that bodies look colored despite not being so, and not because of us. Notice that one could grant this while denying that God brought about E for the purpose of us making false judgments. This would not work if we meant c-judgments, since we will always make c-judgments come what may, and given eliminativism, they will always be false. So this cannot be an objection to D4 per se. In this way, on the intended reading, the objector can grant that D4 is true. However, if she insists that it’s only r-judgments that matter for deception, then the Truth and Function Rules can be re-envisioned simply as guides for what r-judgments God did
intend for \(S\) to form, irrespective of what \(S\) c-judges.\(^{18}\) So while D4 is true for c-judgments, it does not follow that God is a deceiver. Thus D6 is false.

In response, it’s unclear why letting ‘judgment’ refer to c-judgments would render DECEPTION false. If the idea is that c-judgments are not beliefs, our answer is straightforward: c-judgments are belief-like enough. As we have emphasized, c-judgments are issued by the will, and semantically evaluable as true and false; indeed, a judgment that \(p\)—whether a r-judgment or c-judgment—will always involve putting \(p\) before one’s mind as true. If the idea is that c-judgments, at best, show that the sensory system is deceived and not \(S\), then this worry is blocked by the fact that c-judgments are issued by the will—an active, not passive part of the mind-body composite. Thus c-judgments seem perfectly apt to figure into an account of deception.

However, there is certainly something curious happening post-\textit{Meditations}, and perhaps this is what gives rise to suspicions about D6. To make things concrete, consider a subject, \(S\), who has yet to read the \textit{Meditations}. \(S\) has all sorts of false r-judgments: these are the r-judgments that are involved in central planning and conscious inference. Amongst these false r-judgments will be \textit{that bodies are colored}. Like all r-judgments, this is formed from \(S\)’s epistemic habits, and God cannot be impugned by them. But then \(S\) reads the \textit{Meditations}. \(S\) will no longer r-judge that bodies are colored, assuming all goes well. And this will impact \(S\)’s behavior. If you ask \(S\) whether bananas are yellow, or whether \(S\) would bet a large sum of money that the color of healthy grass is green, \(S\) would say ‘no’. So the worry is that, given \(S\)’s new cognitive and behavioral profile, it seems as if \(S\) is now not deceived about bodies being colored, false c-judgments notwithstanding. R-judgments take

\(^{18}\)Pre-\textit{Meditations}, our r-judgments about color are false. Strictly speaking, more than the Truth and Function Rules are needed to show that God is not a deceiver pre-\textit{Meditations}, however. We also need to assume that deception must be intentional. For if there could be deception without an intent to deceive—a matter that is actually contentious, recall fn. 9—one could simply say that we are deceived by God with respect to our r-judgments because we falsely r-judge due to E, even though God only intended for us to form true r-judgments, something we would have known once we read the \textit{Meditations}.\)
priority when assessing the epistemic status of an agent. And you simply cannot be deceived while believing (or r-judging) the truth.

But you can. Take implicit bias. Subjects profess to not have a corresponding explicit belief that \( p \), despite there being evidence that they have the relevant implicit attitude. Likewise, post-Meditations, \( S \) will explicitly deny that bodies are colored, even though her will issues a claim (a c-judgment) to the effect that they are, and even though this issuance is irreversible.

Now it’s often thought that implicit bias is caused by some associative process/associative structure. This is usually glossed as some type of evaluative association, such as an association between a valence (e.g., negative affect) and a concept (e.g. BLACK MALE). Eric Mandelbaum (2016) calls this thesis associative implicit bias (‘AIB’), and convincingly argues that AIB is false. In its place, he proposes The Structured Belief hypothesis. Here, implicit bias is underwritten by unconscious beliefs. These beliefs are not mere associations. They are propositionally structured mental representations to which \( S \) bears the belief relation. So instead of maintaining, to use Mandelbaum’s example, that implicit racists merely associate (say) BLACK MALE and DANGEROUS, the hypothesis is that implicit racists have a belief with the structure BLACK MALES ARE DANGEROUS.

Suppose then that one forms the unconscious belief that black males are dangerous via deceptive means. When asked, you don’t endorse this belief. Yet you repeatedly act in ways that suggest you believe this. For example, when on a subway and a black male sits next to you, you instinctively move away. When you are walking to your car at night, and you see a black male behind you, you quicken your pace. Such examples are emblematic of the behavioral patterns of those who exhibit implicit racial biases. The upshot is that one’s behavior is not merely a function of what one expressly avows, or what one consciously believes.

Whatever c-judgments are—either full-blown beliefs with muted dispositions or belief-like judgments—they are not mere associations. We can
thus substitute Mandelbaum’s implicit beliefs with Descartes’ will-issued c-judgments, and consider our subject once again. S’s false c-judgment is made automatically and is at least confusedly presented to consciousness (Simmons 2003: 554 fn. 13)—much like the implicit racist who unconsciously, and falsely, believes that black males are dangerous. Upon reading the Meditations, S does not simply no longer endorse these judgments. S endorses eliminativism and so believes that bodies are not colored. And this belief is true. Hence S at once c-judges (without endorsing) that p, while r-judging that ¬p, much like classic cases of implicit bias.\footnote{Implicit bias need not be a case of having an unconscious belief that p while refraining from consciously endorsing p; it can also be a case of unconsciously believing that p while consciously believing that ¬p. Bendana and Mandelbaum (forthcoming) call the latter belief fragmentation. This allows for inconsistent information to be selectively accessed such that there is no “interfragment consistency”—that is, no fragment (roughly, a data structure) stores both p and ¬p at once (Ibid.). We contend that post-Meditations the subject is suffering from something like belief fragmentation.} Of course, the implicit racist’s unconscious belief need not be formed via deceptive means. But they can be, and when they are, we have no trouble saying the implicit racist was deceived irrespective of what he or she expressly avows. Her behavior—moving seats on the subway—is then in turn also the by-product of deceptive information. Likewise, S’s c-judgment is occasioned by misrepresentation at the second-grade, and will show up in her behavior, e.g. in how she treats bodies as having the same color across different illumination conditions (supposing we go with Simmons’ story about color constancy). And this all holds irrespective of whether S’s r-judgments are true. As such, the requirement that DECEPTION can only be cashed out in terms of r-judgments is too strong.

3 Don’t Panic

We have argued that Descartes’ God is a deceiver. This may induce panic in the reader. On the face of it, this is a very problematic thesis to attribute to Descartes, for it seems to reveal that Descartes is inconsistent and perhaps
deeply confused. While we think that we have shown that Descartes’ God is a deceiver, regardless of texts that say otherwise, the question before us is whether there is a way for Descartes to accept this conclusion on his own terms. Patterson (2016: 100-102) claims that Descartes cannot. Although we admit that there are no texts where Descartes unambiguously addresses the problem before us, we think that Descartes offers the resources for a solution where we can understand in what sense God can be a deceiver without upending his philosophical system. As such, there is ultimately no reason to panic. The solution isn’t a sensory theodicy by which one tries to explain away the deception. Rather, Descartes has within reach a sensory theodicy in which he can bite the bullet: sensations are deceptive and God is a deceiver on their account, albeit not problematically so.

This task may seem impossible, for there are too many passages where Descartes straightforwardly claims that God cannot be a deceiver. We cannot examine all of these passages here. But consider this paradigmatic one:

It is impossible [non posse] that God should ever deceive me [me unquam fallat]. For in every case of trickery or deception [fallacia vel deception] some imperfection is to be found; and although the ability to deceive appears to be an indication of cleverness or power, the will to deceive is undoubtedly evidence of malice or weakness [malitiam vel imbecillitatem], and so cannot apply to God. (Fourth Meditation, AT VII: 53/CSM II: 37; cf. Principles I.29, AT V.II:16/CSM I: 203; emphasis added).

Here, Descartes claims that every form of deception is incompatible with God’s perfect nature. Why is that the case? The reasoning in this passage

---

20Patterson denies the viability of any Malebranchean solution that says “it is fine for the senses to dispose us to false judgements…since the senses aim at survival rather than truth” (2016: 101). Patterson contends that Descartes can’t go in for this line because he nowhere claims that false c-judgments are designed to preserve our health (Ibid. 105). We disagree: the point isn’t that being false per se is relevant for health-preservation, but that it’s no bar for health-preservation.
(and similar passages) runs as follows. Call this the *incompatibility argument*:

I1 Every form of deception involves malice (one kind of imperfection) or weakness (another kind of imperfection).

I2 Malice is incompatible with God’s benevolence (a divine perfection).

I3 Weakness is incompatible with God’s omnipotence (a divine perfection).

$\therefore$ God cannot be a deceiver.

While premises (I2) and (I3) are implicit in the above passage, the Incompatibility Argument, or something very much like it, represents the central theoretical grounds for denying that God is a deceiver.

However, we *do* argue that the Incompatibility Argument, and the passages that bear it out, do not tell the entire story about Descartes’ views on divine deception. Take (I1). To deny (I1), Descartes would have to allow for:

**Omni-compatible Deception**: God’s $\phi$’ing is an act of omni-compatible deception if God’s $\phi$’ing is an act of deception that (i) does not involve malice (incompatible with omnibenevolence) and (ii) does not involve weakness (incompatible with omnipotence).\(^{21}\)

We think Descartes can allow for omni-compatible deception, and indeed there are passages, if only *glimmers*, that point precisely this way. We’ll focus mostly on clause (i), given that these passages only explicitly reference malice, and because it’s not obvious what Descartes means by ‘weakness’ here anyway. On at least one immediately plausible reading, “malice or weakness” expresses not a true disjunction, but close synonyms. This would make weakness something like *moral weakness*, which is functionally equivalent to malice, thus rendering (I3) and clause (ii) in *Omni-compatible Deception*

\(^{21}\)We set aside features that are incompatible with omniscience, as Descartes does not mention anything related to omniscience in the Fourth Meditation passage above.
superfluous. However, to play it safe, we have tied weakness to omnipotence, with the idea being that surely God could have achieved his ends (whatever they are) in some other way that did not involve deception.\textsuperscript{22} So we’ll have something else to say about this take on weakness as well.

Some will insist that these passages are anomalous, and so worth little weight. But with the arguments of the preceding section, there is now a \textit{meta-interpretive pressure} to look at these passages anew. At a first-order interpretive level—looking at these passages independently from our argument from Sensory Deception—such a dismissal is more warranted, especially in light of Descartes’ denials of God’s deceptive nature elsewhere. However, when these passages are read against the background of our argument, we now have new reasons (e.g. those driving from consistency) to take otherwise obscure passages more seriously.

We proceed in three steps. First, we examine the passages where Descartes discusses the possibility of omni-compatible deception. Second, we turn to maliciousness, and consider the following two questions: (1) is God malicious in giving \( S \) deceptive second-grade color sensations? And (2): Is God malicious in making it such that \( S \) cannot help but issue false c-judgments in response to the second-grade? We’ll argue that the answer to both is \textit{no}. These ‘no’ answers will also allow us to ask \textit{why} God deceives us, which speaks to the question of whether this form of deception implies weakness and thus is at odds with God’s omnipotence. Finally, we argue that omni-compatible deception does not upend other aspects of Descartes’ system.

\section{3.1 Descartes on Omni-Compatible Deception}

There are two main texts where Descartes acknowledges the possibility of divine deception. First, in \textit{Objections and Replies} II Mersenne writes:

\begin{quote}
You say that God cannot lie or deceive. Yet there are some school-
\end{quote}

\textsuperscript{22}Still impressionistically: the thought is that if someone is clever and powerful enough, they would not need to use deception to get what they want.
men who say he can. Gabriel...and Ariminensis among others, think that in the absolute sense God does lie \[mentiri\], that is, communicate to men things which are opposed to his intentions and decrees. Thus he unconditionally said to the people of Nin- eveh, through the prophet, ‘Yet Forty days and Nineveh shall be destroyed.’ And he said many other things which certainly did not occur, because he did not mean his words to correspond to his intentions or decrees...Cannot God treat men as a doctor treats the sick, or a father his children? In both these cases there is frequent deception \[saepe decipiunt\] though it is always employed beneficially and with wisdom \[sapientur & cum utilitate\]. For if God were to show us the pure truth, what eye, what mental vision, could endure it? (AT VII: 125-6/CSM II: 90).

In pointing out false texts in scripture that may lead to false judgments, Mersenne is pressing on Descartes’ seemingly unrestricted contention that God cannot be a deceiver. Mersenne offers the doctor-patient example to help explain how deception could be permissible for God. A doctor might say something false to a patient about their sickness with the intent of helping them. But this isn’t obviously problematic. Similarly, there might be falsities in scripture that are beneficial for believers. Perhaps then such cases of deception would be permissible for God because it doesn’t conflict with his benevolence. Now notice how Descartes responds:

\[Iw o u l dn o tw a n tt oc r i t i c i z e\] \[nolim tamen reprehendere\] those who allow that through the mouths of the prophets God can produce verbal untruths which, like the lies of doctors who deceive their patients in order to cure them, are free of any malicious intent to deceive \[desit omnis malitia deceptionis\]. (AT VII:

\[23\] Note that \[mentiri\] is translated by Cottingham et. al. as “lie.” But a more suitable translation is “deceive” given the broader context of this passage. For a discussion of the medieval accounts of deception referred to here, see Perler (2010).
While Descartes does not explicitly endorse non-malicious deception, that he claims that he will not criticize the possibility of the doctor-patient case suggests an implicit endorsement. Throughout the *Objections and Replies*, Descartes is quite happy to point out when his interlocutors fall into contradictory positions. If it were in fact contradictory for God to deceive full stop, then Descartes would presumably point that out. One part of his response is particularly telling: he claims that doctors can deceive without having “any malicious intent to deceive.” This stands as a qualification to (II).

Second, Hobbes raises the same query:

> The standard view is that doctors are not at fault if they deceive their patients for their health’s sake, and that fathers are not at fault if they deceive their children for their own good. For the crime of deception consists not in the falsity of what is said but in the harm [*injuria*] done by the deceiver. (*Objections III*, AT VII: 195/CSM II: 136)

Descartes responds like this:

> [I don’t] require that we can in no case be deceived (indeed, I have readily admitted that we are often deceived). All that I require is that we are not deceived in cases where our going wrong [*noster error*] would suggest an intention to deceive on the part of God [*voluntatem in Deo testaretur*]; for it is self-contradictory that God should have such an intention. (*Third Replies*, AT VII: 195/CSM II: 136-7)

This response is especially interesting, if only because Descartes doesn’t answer Hobbe’s question. The doctor’s main intention is to aid the patient; the deception is a means to that end. So, again, we wonder: if this were something
Descartes wanted to definitively rule out, why did he not say so?24

Descartes is cagey in these two texts. He is of course well-aware of all the places where he claims that God cannot be a deceiver, and the foundational role that this claim plays in his metaphysics and epistemology. Descartes might have feared, then, that he would upend his entire system, and thus did not want to make a full-blown concession to Mersenne and Hobbes. Yet our contention is that Descartes failed to realize two crucial things: first, that his prior commitments concerning the structure of our sensory systems commits him to God being a deceiver in some sense, and second, that in fact he can admit this without wreaking havoc on his philosophical system.

Suspending judgment regarding whether this is what Descartes actually believed or meant, here is our suggestion for what Descartes could—and indeed in light of our preceding argument from Sensory Deception—should have meant: some acts of deception are not malicious. That is, they are not meant to, nor cause, harm. And thus such acts of deception are omni-compatible insofar as they are compatible with God’s benevolence.

Equating non-maliciousness with non-harmfulness is sensible given the standard meaning of ‘malice’ but also Hobbes’ point that the “crime of deception consists...in [its] harm.” Harm [nociturus] is often bodily, as is the case in Descartes’ Sixth Meditation discussion of dropsy (AT VII: 85/CSM II: 59). But other harms are non-bodily. In the case of the evil deceiver of the First Meditation, the harm is plausibly psychological (AT VII: 23/CSM I: 15). This is evidenced by the kinds of passions—such as dread and fear—expressed by the meditator when she considers the possibility that she could be deceived about all matters (see Section 3.3.3).

Now there are two components of deception that are relevant for present

24 Other commentators have read Descartes as conceding to Hobbes the possibility of omni-compatible deception. For example, Robinson writes: “According to Descartes, Mersenne and Hobbes are right; there are cases of morally acceptable deception. When, in the Meditations, Descartes says that it is impossible for God to deceive, we should not understand him as maintaining that God cannot deceive simpliciter. Rather, he should be understood as claiming only that God cannot deceive maliciously” (2013: 522).
concerns. First, there are the *deceptive* appearances at the second-grade, i.e. sensory evidence E. Second, there are the false c-judgments that we issue in response to E. Let God be a malicious deceiver \(\text{iff} \) either of these components harms the subject who is deceived. So, the question we must now answer is this: is the deception inherent in our sensory systems malicious?

### 3.2 Benefit and Harm in Sensory Systems

Ascertaining whether God’s designing of E makes him malicious can be clarified by considering the following two claims:

**Biological Function:** The biological function of sensations is to preserve the health of the mind-body composite. (Sixth Meditation, AT VII: 83/CSM II: 57)

**Best Sensations:** God gave us the best of all possible sensations. (Sixth Mediation, AT VII: 87-8/CSM II: 60)

The conjunction of Biological Function and Best Sensations implies that God gave us the best sensations *qua* their functional role—“the one sensation which, of all possible sensations [\textit{ex omnibus quos inferre potest}], is most especially and most frequently conducive to the preservation of the healthy man” (Sixth Meditation, AT VII: 87-8/CSM II: 60). So the tighter gloss on the current question is this: assuming God would be malicious if he gave us anything less than the best of all possible sensations, is the conjunction of Biological Function and Best Sensations compatible with God giving us systematically misleading color sensations?

In response, we won’t defend the claim that the misleading nature of color sensations somehow makes them ‘the best’ *qua* their aiding us in (doing things like) finding food sources. To maintain compatibility, color sensations being misleading simply need be \textit{no bar} to their optimally playing this sort of role.
When asked how our various sensations could play their part in the preservation of the mind-body composite, whether they misrepresent is irrelevant. What matters is reliability. Mendelovici (2013) points out that a test can fail to be valid while still being reliable. For example, the SAT may fail to be valid insofar as it fails to detect what it’s intended to detect—e.g. predicting success in college (Ibid.: 422). Still, it could be reliable insofar it tends to yield the same results in different contexts. Similarly, while some representations might be veridical and reliable, others may not. Our interest is in those that reliably misrepresent: reliable, in that they respond similarly in similar circumstances; and non-veridical, since the world does not have the properties they represent it to have. We want, as Mendelovici puts it, a representation that “[gets] things wrong in the same way all the time” (Ibid.: 423).

Color sensations are reliable misrepresentations. They are illusory, but in a systematic and reliable way. Paraphrasing somewhat, Mendelovici (Ibid.: 423) tells us that, for any (indicative) mental representation R and any property F, R reliably misrepresents F iff: (i) R represents objects as having F; (ii) Most or all of the relevant objects don’t have F; and (iii) Tokens of R do or would non-veridically represent objects as having F in the same types of circumstances on separate occasions. These conditions are sufficient to reconcile God’s designing us to have systematically misleading color sensations with Biological Function and Best Sensations, and thus compatible with God being a non-malicious deceiver. Here’s why. Suppose S only wants to identify ripe bananas. Ripe bananas look yellow. By eliminativism, they are not yellow. So, when S sees bananas, she represents them as having a property they don’t have. But that’s not a problem insofar as finding food goes. What would be a problem is if S’s sensation represented the banana as yellow in context C, green in context C*, magenta in C**, and so on, where C, C*, and C** are all tokens of the same context-type. In all contexts, S’s sensation would misrepresent. But it wouldn’t be doing so in a reliable manner. And that would prevent S from finding food in an optimal manner, since
it would sever the connection between *being ripe* and *looking yellow*.

Considerations that hold at the second-grade can be extended to the third-grade. C-judgments are part of our sensory system, whose function is the preservation of our health. Just as the non-veridicality of color sensations is irrelevant on this score, the falsity of c-judgments is too. What matters is the properties they represent bodies as having, and how they do this.

That said, in the Fourth Meditation, Descartes tells us that false judgments are a *privation*, and that God cannot be causally responsible for any privation (AT VII: 54-5/CSM II: 38). A privation is the lack of a perfection that a substance ought to have (Ibid.). In this way, privations are malicious, because if a substance lacks a perfection that it ought to have, it’s harmed. Thus perhaps a further worry remains.

One move invokes the standard Cartesian response: the source of the privation of false judgments is $S$, not God, because $S$ forms false judgments through a *misuse* of the will, and can avoid these false judgments by adhering to epistemic norms like the *Truth Rule*. But of course, as we have argued, that won’t work here; at the third-grade of sensory response, $S$ cannot help but form false c-judgments, and there are no epistemic norms that could get $S$ out of these false c-judgments. This would seem to make God causally responsible for c-judgments, and therefore malicious.

Yet the worry trades on the assumption that false c-judgments are a privation to begin with. We reject this. False c-judgments are not detrimental to the health of the mind-body composite. Yet since a false c-judgment would be a privation only if tokening that false c-judgment *would* be detrimental to our health, it follows that false c-judgments are not privations.

This necessary condition on ‘privation-hood’ is motivated by two points. First, Descartes’ claim about privations in the Fourth Meditation needs to be indexed to the epistemic and metaphysical progress the meditator has made with respect to her understanding of her *self*. By the Fourth Meditation, the meditator understands herself solely as a thinking thing, and in the
context of that meditation, the task is to establish the standards for *scientia*. However, by the Sixth Meditation, the body also belongs to *S*, in some sense (Chamberlain 2019b). *S*’s sense of self has expanded, thus our understanding of a privation, in the context of the mind-body composite—ought to shift as well. Given that *S* has a body whose health must be preserved, a sensory privation in this context is just a sensory response that is harmful to the mind-body composite. And it turns out that false c-judgments are a sensory response that is beneficial. Therefore, they are not a privation.

Now notice: in the current story, deception *per se* need not be beneficial. In the patient-doctor example, what is benefiting the patient is the doctor’s *speech act*—her utterance of ‘you are going to get better.’ That claim might be false, and the doctor might know it. Since the doctor wants her patient to believe that she will get better, the doctor deceives her patient. Still, the speech act isn’t beneficial in virtue of the statement it expresses being false. After all, the speech act would *still* be beneficial even if it were true. When a doctor says to a patient ‘you are going to get better,’ the reason this helps (or so we are told) is that it aids the patient psychologically. These effects would still result even if it were true that the patient was going to get better. (Maybe the doctor was right to say this even though she didn’t know it.) Similarly, color sensations are not beneficial in virtue of being deceptive, but in virtue of the way they attribute colors to bodies. They are deceptive, as bodies don’t have colors. The deception is simply non-malicious.

This point does, however, harken back to the weakness component of *Omni-compatible Deception*. For now we are left wondering why God deceives us, even if only non-maliciously? If deception *per se* isn’t beneficial as just conceded, it’s natural to wonder why God didn’t simply make bodies colored, and let us reap the benefit anyway. That might strike some as a sign of weakness, which is certainly not a concession Descartes can make.

The cogency of this question assumes that God *could* have made color a property of *res extensa* without making color sensations any less beneficial.
Yet this assumption is false. There is no sense in which God ‘could’ have created a world in which bodies were colored because Descartes does not, like Leibniz, have a modal metaphysics. Descartes’ metaphysics of God rules out any substantive sense of possible worlds and unactualized possibilities (Nelson & Cuming 1999). God is deeply metaphysically simple: His attributes are identical to each other (Principles I.23 AT VIIIA: 12-4/CSM I: 200-1). Most relevant for present purposes, if God’s intellect and will are identical, this implies that God’s thinking is identical to his creating. If there were unactualized possible worlds in Descartes’ metaphysics, they would have to exist in God’s intellect. However, the identification of intellect and will implies that God creates whatever he considers (Ibid.: 139).

In addition, this assumption presupposes a backwards explanatory order: that we can theorize about the nature of our sensory systems independent from what bodies are actually like. This too is false. Descartes’ metaphysics here is world first. Since disembodied souls don’t sense anything—they have pure intellection—the very question of the nature of sensation, and what kind of sensory system would be best, becomes live only once souls are embodied, and put into an environment occupied by other bodies. Moreover, what kind of sensory system would be best is fixed by the nature of that environment, given what Descartes says about the function of sensations. So the question ‘why is the nature of bodies such-and-such?’ is prior to the question ‘why do we have such-and-such sensations?’ The latter arises only once answers to the former are fixed. This suggests a methodology when thinking about the nature of our mental lives qua embodied beings, viz. that we must seek explanations against background assumptions concerning the nature of res extensa (cf. Carriero 2009: 423). If this is correct, and one concurs that bodies are not colored, then all we need to do is ask what would be better: a sensory system that represents bodies as colored (and is thus deceptive) or one that doesn’t (and thus isn’t deceptive, at least in that respect). Clearly the answer is the former. Sensory deception is thus compatible with God’s
being all-powerful, and thus is omni-compatible on that count.

3.3 Systematic Concerns

The veridicality of clear and distinct perceptions, the resolution of the dream doubt, and the proof of the external world rely on some anti-deception premise. Yet all that’s needed is an anti-malicious deception premise.

3.3.1 The Veridicality of Clear and Distinct Perceptions

A foundational claim of the *Meditations* is that clear and distinct perceptions are veridical because God is not a deceiver. Could God non-maliciously deceive us about clear and distinct perceptions?

No. If God were to deceive $S$ about her clear and distinct perceptions, this would be malicious deception, which is exactly why Descartes says that it’s inconsistent with God’s nature for him to deceive us about our clear and distinct perceptions. One reason intellectual deception would be malicious is that it would prevent $S$ from attaining virtue and ultimately happiness. And for God to prevent $S$ from attaining happiness would be to harm her. For Descartes, genuine happiness—namely, the supreme satisfaction or contentment of mind—results from being virtuous (see, e.g., Letter to Queen Christina 20 Nov. 1647, AT V: 83/CSMK: 325). Cartesian virtue consists in the firm and constant resolution to use the will well. Here, using the will well consists in practically judging and acting according to knowledge of the truth (Letter to Princess Elizabeth 15 Sept. 1645, AT IV: 291/CSMK: 266). Knowledge of the truth is a necessary (but not sufficient) condition for virtue, and broadly construed consists of four truths: the existence of God, the distinction between mind and body, the immensity of the universe, and the interconnectedness of the universe (Letter to Princess Elizabeth 15 September 1645, AT IV: 291-3/CSMK: 265-6). The first two truths are metaphysical, the latter two fall under natural and moral philosophy. Let us focus on the
first two truths. According to Descartes, $S$ needs to know that God exists and that everything in existence is a manifestation of God’s will, and $S$ must also know that the mind is distinct and more noble than the body in order to regulate our passions and exercise virtue (Ibid.). These two metaphysical truths are grounded in clear and distinct perceptions (*redacted*).

If $S$ did not know these truths, Descartes would say it’s impossible for $S$ to exercise virtue, and thus she would be incapable of acquiring virtue and happiness, at least in their ideal forms (Dedicatory Letter to Elizabeth, AT VIII A: 3/CSM I: 191). This implies that if God were to deceive $S$ about say, the real distinction between mind and body, then He would be preventing $S$ from acquiring ideal virtue and happiness—a clear harm. As such, intellectual deception is inconsistent with God’s benevolent nature.

### 3.3.2 The Dream Doubt

Descartes’ resolution of the dream doubt also depends on the claim that God is not a deceiver. The dream doubt in the First Meditation is that $S$ has no reliable way of distinguishing waking states from dream states, and is specifically designed to cast doubt on our certainty in the existence of external bodies. At the end of the Sixth Meditation, Descartes claims that $S$ can distinguish between waking states and dream states: “I now notice that there is a vast difference between the two, in that dreams are never linked by memory with all the other actions of life as waking experiences are” (Sixth Meditation, AT VII: 89/CSM II: 61). Since God is not a deceiver, and $S$ now knows that memory is reliable, I can rely on reports of memory (as well as other faculties) to verify whether I am in a waking state.

That God is a non-malicious deceiver causes no problem here. For if God were to deceive $S$ about the *existence* of bodies, he *would* be harming $S$, for she would be out of touch with reality, which would cause psychological harm.\(^{25}\) The psychological harm induced by not being in touch with reality is

\(^{25}\)Notice that if the dream doubt cannot be resolved, then we risk not being in touch with
brought out at the end of the First Meditation, where the meditator expresses “dread” and “fear” in the face of the “inextricable darkness of the problems I have now raised” (First Meditation, AT VII: 23/CSM II: 15). If Descartes could not resolve the dream doubt, much of these passions and thoughts would persist, yielding a psychologically and practically harmful state of affairs, preventing S from living a psychologically healthy life.

3.3.3 The Proof of the External World

Descartes’ proof of the external world is handled in the same way. The proof is a disjunctive syllogism (Stuart 1986: 22). Either my ideas of material objects are caused by (i) some unknown faculty of the mind, (ii) God, or (iii) by actual material objects. Yet my ideas of material objects are neither caused by (i) some unknown faculty of the mind, nor (ii) by God. Hence, they must be caused by (iii) actual material objects. Now the sub-argument against (ii) appeals to God’s non-deceptiveness (Sixth Meditation, AT VII: 79-80/CSM II: 55), but clearly all that’s relevant here when it comes to the argument’s validity is God’s non-deceptiveness with respect to the existence of external bodies. Yet as we have seen, perceiving us in that respect is malicious.

4 Conclusion

Descartes is committed to Sensory Deception. This thesis, plus facts about our sensory system, entails that God is a deceiver. But when properly understood in terms of non-malicious deception, Descartes can admit all of this. His broader system remains intact. Yet we are required to rethink how to approach that system.

reality at all. Of course, there is a sense in which color sensations leave us out of touch with reality, but they nonetheless will accurately represent other properties of bodies besides colors. Color sensations are illusions, and it’s a conceptual truth that illusions (unlike hallucinations) make us aware of the objects they represent.
The line ‘God is not a deceiver’ is canonical as the theoretical linchpin of the *Meditations*. We don’t deny that presentational simplifications are often permissible if not advisable, especially for the uninitiated, as when teaching courses that serve to introduce philosophy in general or early modern philosophy in particular. But not all simplifications are equal, and in the present case, not only is the line false—in fact, God *is* a deceiver—once we take the aforementioned lesson to heart, we see that deception *per se* need not be at the foundation of Descartes’ system.

The *Meditations* contains two theodicies: the epistemic theodicy of the Fourth Meditation and the sensory theodicy of the Sixth Meditation. In the Fourth Meditation, Descartes needs to resolve the epistemic worry that God might deceive us in our clear and distinct perceptions, which would preclude us from acquiring *scientia*. Having completed the Fourth Meditation, it’s tempting to attribute to Descartes the view that God can never be responsible for our forming *any* kind of false judgments. The problem is that this makes a consistent sensory theodicy in the Sixth Meditation impossible. For in the *Sixth Replies*, Descartes clarifies his account of our sensory systems so as to make clear that they include c-judgments: will-issued, irrevisable, false sensory judgments. However, if we see Descartes as being primarily concerned with *benevolence*, the tension dissipates. For now the Fourth Meditation in fact teaches us that God cannot be responsible for us forming any *harmful* judgments. That’s bedrock. And since *false* r-judgments about clear and distinct perceptions *would* be harmful, they cannot be due to God. However, the standards for harm in the sensory theodicy of the Sixth Meditation are different. C-judgments are not harmful on account of their being false, but they are beneficial for our health in virtue of the way they represent bodies. So God, in causing us to have c-judgments, is both a deceiver *and* benevolent. What’s ironic is that it takes realizing that God is a deceiver to properly see the theoretical centrality afforded to benevolence within Descartes’ system.
References


[31] Nolan, L (2011). Descartes on “What we call color”. In *Primary and Secondary Qualities: The Historical and Ongoing Debate*. OUP.


