

LOCKE AND DESCARTES ON MENTAL TRANSPARENCY

Philipp N. MÜLLER*

Abstract. The transparency thesis – i.e. the doctrine that every mental state is necessarily conscious – was a widespread view in early modern philosophy. In this paper, I inquire into the role of mental transparency in the philosophies of John Locke and René Descartes. I begin by sketching a shared Lockean-Cartesian picture of mind as it pertains to the psychological or structural aspects of consciousness. I then distinguish mental transparency from the closely related concept of epistemic transparency and argue that the thesis must allow for different degrees of conscious awareness, which is needed to address some of our uneasy intuitions. Afterwards, I examine Locke’s and Descartes’s reasons for adopting transparency in their respective philosophies. In the case of Descartes, I present consciousness as a necessary condition for knowledge of our own minds in the larger context of his epistemological goals in the *Meditations*. In the case of Locke, I examine three of his arguments in order to illustrate the indispensable role of transparency in his polemic against central Cartesian doctrines such as innatism and the thesis that the soul always thinks.

Keywords: Consciousness, John Locke, memory, René Descartes, transparency

1. Introduction¹

The transparency thesis – i.e. the doctrine that every mental state is necessarily conscious – was a widespread view in early modern philosophy.² To our modern sensibilities, however, this seems surprising, given that we conduct much of our daily lives on the assumption that our minds do all sorts of things we are not aware of. Our sense of ourselves and of others appears to be closely intertwined with considerations of unconscious wishes, desires, and beliefs. It therefore seems plausible to conclude that the mind engages in all sorts of unconscious or subconscious perceiving, processing, and – ultimately – thinking. This is not a simple case of our everyday intuitions and the harsh reality of academic inquiry coming apart. As Alison Simmons puts it:

* Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany.
philipp.n.mueller@posteo.de.

Today such a claim seems either hopelessly naïve or blindly dogmatic, and certainly wrong. Empirical work in cognitive and social psychology suggests that so much of our mental life trundles along unconsciously it is a wonder the mind bothers with consciousness at all.³

The main target of such rhetoric, of course, is René Descartes, and the term “Cartesianism” has become synonymous with the thesis of mental transparency. But Descartes was not its only proponent; in fact, the transparency thesis would later be endorsed by some of the most prolific thinkers of the seventeenth and eighteenth centuries, including Louis de La Forge, Nicolas Malebranche, and – most importantly for our purposes – John Locke. But the question remains: Why did they adopt a thesis that has such a counterintuitive ring to it?

In this paper, I will propose an answer to this question, at least with regard to Locke and Descartes. To do so, I will take a closer look at their shared commitment to mental transparency, especially the important use of it by Locke in his polemic against cherished Cartesian doctrines. In the first section, I will sketch a shared Lockean-Cartesian picture of mind as it pertains to psychological or structural aspects of consciousness. This will not serve simply to point out the irony of Locke’s instrumentalization of their shared assumption against his intellectual forerunner but is meant rather to illustrate the distinct metaphysical conclusions Locke is able to draw from the remarkable structural similarities between Descartes’s theory of mind and his own. In the second section, I will turn to the thesis itself. I will first distinguish it from the closely related concept of *epistemic* transparency; then, in order to address some of our uneasy intuitions, I will argue that the thesis must allow for different degrees of conscious awareness – i.e. that although we are conscious of all our thoughts, this does not imply that we are *equally* conscious of all our thoughts; finally, I will examine Locke’s and Descartes’s reasons for incorporating this into their respective philosophies of mind. The third section will then show how the transparency thesis plays a central role in Locke’s broader anti-Cartesian polemic, especially in furnishing him with arguments against Cartesian metaphysics of mind; three of these arguments will be examined.

2. Lockean Consciousness in the Cartesian Tradition

In this section, I will outline the common ground shared by Locke and Descartes concerning the mental. Let us start with what both consider to be the principal objects of consciousness: ideas. Both Locke and Descartes use the term “idea” to cover a very wide array of basic mental items. Already in the Introduction to the *Essay*, Locke apologizes for his frequent use of the word:

It [i.e. “idea”] being that Term, which, I think, serves best to stand for whatsoever is the Object of the Understanding when a Man thinks, I have used it to express whatever is meant by *Phantasm*, *Notion*, *Species*, or whatever it is which the Mind can be employed about in thinking; and I could not avoid frequently using it.⁴

Descartes admits to comparably broad use of the term in his *Reply* to Gassendi: “[Y]ou restrict the term ‘idea’ to images depicted in the imagination, whereas I extend it to cover any object of thought.”⁵ It would seem that both philosophers build much of their respective philosophies of mind on ideas as the most basic mental unit. When accompanied by conscious perception, ideas form the contents of thoughts, insofar as they are what we are conscious of when thinking. Without ideas as the objects of consciousness, there would be nothing to be aware of and no thinking could occur. In the sense that they account for everything we can be cognizant of, they are what enables us to be aware not only of the world, but also of ourselves.⁶

But how does consciousness relate structurally to its objects? More specifically, how is the act of perception that enables us to be consciously aware of our mental goings-on tied to its contents? I would like to briefly advance the thesis that it is possible to trace a unified conception of consciousness, already present in Descartes, across successively more developed forms expounded by his followers, all the way to its most mature and explicit form in Locke.⁷ On this view, consciousness is a reflexive, perceptual awareness internal to every thought, by virtue of which we come to be aware of the thought’s contents; this conception is thus a first-order theory of consciousness. It should be stressed, however, that I claim neither that Locke’s and Descartes’s respective philosophies of mind mirror each other exactly, nor that Locke simply followed the Cartesian view, but rather that central aspects of Locke’s views were heavily influenced by and informed by Descartes. In a way, the picture of mind found in the *Essay* is part of his Cartesian inheritance, i.e. the most mature form of a tradition that originated with Descartes and was developed further by his followers, such as La Forge and Arnauld, as will be shown below.⁸

There is some textual evidence in Descartes’s work that speaks in favor of a first-order reading. Consider for example the passage in the Second Set of Replies where Descartes says that ideas are “the form of any given thought, *immediate perception* of which makes me aware of the thought.”⁹ A similarly close connection between thoughts and our consciousness of them can be found in his Third Set of Replies (to Hobbes): “[W]hen I want something, or am afraid of something, I *simultaneously* perceive that I want, or am afraid.”¹⁰ And finally, in *The Passions of the Soul*, he argues that “it is certain that we cannot will anything without thereby perceiving that we are willing it. But [...] this perception is really one and the same thing as the volition.”¹¹ Descartes seems to be saying that conscious awareness of our thoughts amounts to perceiving them, and moreover that this perception is (i) immediate (as per the Second Set of Replies), (ii) simultaneous (as per the Third Set of Replies), and (iii) somehow “built into” the thought in question (as per the *Passions*). From (i) and (ii) we can draw the conclusion that there is no “gap,” as it were, between a thought and the consciousness of that thought – which (however small) would arguably have to be the case if consciousness were a higher-order phenomenon, i.e. if there were another, distinct thought responsible for conferring consciousness on the lower-order thought in question. Moreover, (iii) strongly implies that there is no numerical distinction between a thought and that which makes it conscious. This selection of textual

evidence suggests that the perception that accounts for consciousness is inherent in the thought and is therefore not a higher-order phenomenon.¹²

In their development of Descartes's thought, his followers were quite explicit about their first-order views on consciousness. Louis de La Forge, for example, in his *Traité de l'esprit de l'homme* not only states clearly that the perception relevant for consciousness is internal to thought, but also draws a distinction between consciousness and the kind of explicit, reflective, second-order awareness of our own mental goings-on that we later find in Locke, who calls it simply "reflection":

I think I can define the nature of thought as that consciousness, awareness and inner feeling by which the mind is aware of everything it does or suffers and, in general, of everything which takes place immediately in itself at the same time as it acts or is acted on. I say "immediately" to let you know that this testimony and inner feeling is not distinct from the action or passion and that the actions and passions themselves make the mind aware of what is taking place in itself. Thus, you will not confuse this inner feeling with the reflection that we sometimes make on our actions, which is not found in all our thoughts because it is only one type of thought.¹³

While there are certainly original elements in La Forge's development of Descartes's thought, the way he makes his points is similar to what Descartes says in the passages examined above. Other potential similarities between Descartes, La Forge, and Locke also emerge from this passage. Villi Lähtenmäki argues persuasively that a distinction between consciousness and reflection is already implicitly present in Descartes.¹⁴ To do so, he distinguishes three kinds of consciousness: (i) rudimentary consciousness, (ii) reflexive consciousness, and (iii) consciousness achieved by attentive reflection. Note that (i) and (iii) seem to parallel the distinction made by Locke and La Forge between consciousness and reflection. Though Lähtenmäki's objective is to resolve the tension between conflicting passages within Descartes's works supporting either a first-order or a higher-order view of consciousness, the fact that both concepts are already present in his philosophy, at least to some degree, only speaks in favor of my general point that Locke's views on consciousness are informed by and influenced by the Cartesian tradition.

Both the view of consciousness as a first-order phenomenon and the distinction between consciousness and reflection can in turn be found in an even more mature form in the works of Antoine Arnauld, with whom Locke was intimately familiar. Most notably, in *Des vraies et des fausses idées* Arnauld unambiguously states that "thought or perception is essentially reflective on itself, or, as it is said more aptly in Latin, *est sui conscia*. For I do not think without knowing that I think."¹⁵ Here, consciousness is once again characterized as a reflexive (and therefore internal) property of thoughts, which he later calls *réflexion virtuelle*, to distinguish it from *réflexion expresse*, which is just the sort of attentive, higher-order introspection that is, on Lähtenmäki's reading, already implicitly present in Descartes¹⁶ and later addressed explicitly by La Forge, and which would later appear as Locke's concept of reflection.

What then of Locke's own remarks on consciousness proper (as opposed to reflection)? While there is still some debate on the issue as it pertains to Descartes, it appears that commentators now generally agree that Locke too was a first-order theorist of consciousness.¹⁷ Let us briefly survey the textual evidence. Consider the following passage, which immediately precedes the infamously ambiguous assertion that "consciousness is the perception of what passes in a man's own mind":

If they say, The Man thinks always, but is not always conscious of it; they may as well say, His body is extended, without having parts. For 'tis altogether as intelligible to say, that a body is extended without parts, as that any thing *thinks without being conscious of it*, or perceiving, that it does so. They who talk thus, may, with as much reason [...] say, That a Man is always hungry, but that he does not always feel it: Whereas hunger consists in that very sensation, as thinking consists in being conscious that one thinks.¹⁸

Locke's use of the word "consist" to explain the relationship between a thought and the consciousness of that thought suggests an internal relation between the two. On this reading, consciousness is an *intrinsic* property of mental states in much the same way as it is for the Cartesians. Furthermore, Locke's analogy with bodies implies that this relation is non-contingent, and thus necessary. On the higher-order account, by contrast, consciousness would be a phenomenon *extrinsic* to mental states, since any given mental state is conscious only by virtue of being represented by another state. And since, on this view, we would be dealing with two numerically distinct states – a lower-order thought that is made conscious and a higher-order thought that makes the lower-order thought conscious – which could come apart from one another, the relationship between them could never be anything but metaphysically contingent.¹⁹ Even section II.xxvii.9 of the *Essay* – the passage where Locke explains that it is "impossible for any one to perceive, without perceiving, that he does perceive" – is qualified by the immediately preceding assertion that consciousness "is *inseparable* from thinking, and as it seems to me essential to it."²⁰ Again, Locke's choice of words clearly suggests a theory of consciousness according to which the relationship between our mental states and our awareness of them is not that of representation by a distinct, contingently present mental state. What he seems to be saying instead is that the "perception of perception" that constitutes consciousness is not a higher-order state, but a reflexive, internal component of the original perception itself, and therefore a phenomenon on the same level.

Lastly, there is a solid systematic reason for embracing a first-order reading of both Locke and Descartes, one that is related to the main topic of this paper – namely, the transparency thesis. As was originally pointed out by Aristotle in *De anima* 3.2 and later brought to bear against Locke by Leibniz,²¹ holding a higher-order account in conjunction with the transparency thesis will result in an infinite regress: for if all thinking is conscious, and if every conscious thought is accompanied by another, higher-order thought, there will be an infinite regress of thoughts having other thoughts as their object. Thus, if Locke and Descartes embraced a higher-order

account of consciousness, then their accounts would be obviously defective. So stark a clash with a commitment as central and as frequently stated as mental transparency is best avoided, at least from a systematic point of view.²²

3. Mental Transparency

What has been explained above amounts to a rough description of a working theory of consciousness for both philosophers. But the view that consciousness is due to a reflexive perception internal to thought does not by itself entail that all thinking is conscious. While the textual evidence cited in favor of this interpretation is already closely intertwined with talk of mental transparency, these two aspects of consciousness are conceptually distinct. Indeed, it would be entirely coherent to hold that while we do have first-order consciousness of those thoughts we end up being aware of, we simply do not end up being aware of *all* our thoughts. We must therefore ask why Locke and Descartes felt the need to posit that all thinking is conscious. Answering this question will be my overarching goal in the rest of this paper.

But we should perhaps first clarify the question a bit more. Both Descartes and Locke are unambiguous about their commitment to the doctrine that all thinking is conscious. In his *Reply* to Arnauld, Descartes writes:

For there is nothing that we can understand to be in the mind, regarded this way, that is not a thought or dependent on a thought. [...] [W]e cannot have any thought of which we are not aware at the very moment when it is in us.²³

Similarly, Locke repeatedly declares his staunch adherence to the thesis – for example, when he proclaims that the soul “*must necessarily be conscious of its own Perceptions.*”²⁴ So far, the transparency thesis is as easily stated as it is bold: every thought is necessarily conscious. But as it stands it also holds potential for ambiguity: for if we are *conscious* of all our thoughts, one could easily be led to conclude that we also have some kind of *knowledge* of all our thoughts. Don’t we know what is going on in our minds by virtue of being conscious of its contents? The concept of epistemic privilege concerning the mental – i.e. the view that we have a special kind of first-person, incorrigible knowledge of our own thoughts – not only is closely related to mental transparency, but is also present in both Locke and Descartes.²⁵ In order to get a better grasp of transparency, we should now turn to getting rid of this ambiguity by pulling apart the concepts of epistemic privilege and mental transparency.

For both thinkers, it is certain that consciousness is closely connected with a distinct epistemic privilege and incorrigibility of the subject when it comes to its own mental affairs. On this view, we cannot be mistaken about the contents of our own thoughts, i.e. the ideas present to our minds. The Cartesian meditator confesses to himself: “Yet I certainly *seem* to see, to hear, and to be warmed. This cannot be false.”²⁶ This claim is strikingly paralleled in Locke’s *Essay*: “When we see, hear, smell, taste, feel, meditate, or will any thing, we know that we do so.”²⁷ But what is the exact connection between mental transparency (the thesis that all thoughts are necessarily conscious), and epistemic incorrigibility (the thesis that we have certain knowledge of

our thoughts)? If we accept one of these theses, are we perhaps even forced to accept the other?

Conceptually speaking, the answer to the latter question is no. Though the two theses seem at first glance to be closely connected, each of them requires at least one additional premise to imply the other. Incorrigeability, on the one hand, entails transparency only if the notion is taken to rule out ignorance, as opposed to mere error. For it would be entirely coherent to suppose that while we do have certain knowledge of our mental affairs, this knowledge still might extend only to those thoughts we are cognizant of in the first place. We cannot be guilty of epistemic error concerning a given thought in cases where we are simply ignorant of it. Transparency, on the other hand, precludes the subject having epistemically “opaque” thoughts only if we add the further premise that *consciousness* of one’s mental affairs directly entails *knowledge* of them.

A question that then naturally arises is: How are these two theses linked to each other in the philosophies of Descartes and Locke specifically? Answering this question will require further consideration of their respective accounts of knowledge. In the case of Descartes, there is some degree of *prima facie* textual evidence suggesting that he adheres to epistemic incorrigibility in the *Meditations*. The indubitability of the meditator’s “I think” is central to the Second Meditation, and this notion of thinking is then expanded to cover a variety of mental states: “But what then am I? A thing that thinks. What is that? A thing that doubts, understands, affirms, denies, is willing, is unwilling, and also imagines and has sensory perceptions.”²⁸ The meditator’s rhetoric suggests that the knowledge acquired in introspection is so far beyond doubt that it can serve as the foundation for everything that follows in his later considerations. Initially, therefore, it appears that everything that the light of consciousness touches is subject to indubitable, certain knowledge.

But accepting this claim is not entirely unproblematic. For as several commentators have pointed out,²⁹ there are significant tensions between it and various passages scattered across Descartes’s body of work where he clearly accepts the possibility of error about occurrent thoughts. This is due to the mental act accounting for consciousness and the mental act accounting for knowledge being distinct from one another, as is clear, for example, when he says that “believing something and knowing that one believes it are different acts of thinking, and the one often occurs without the other.”³⁰ To have knowledge in Descartes’s sense, we need to form judgments about our thinking, which may be subject to error. Consider the meditator’s famed reflections on a piece of wax:³¹

[T]he perception I have of it is a case not of vision or touch or imagination – nor has it ever been, despite previous appearances – but of purely mental scrutiny; and this can be imperfect and confused, as it was before, or clear and distinct as it is now, depending on how carefully I concentrate on what the wax consists in.³²

He is then “amazed at how weak and prone to error [his] mind is.” What is important here is that the meditator’s initial conscious perception of the wax as defined by its primary qualities was imperfect and confused. He then discovers that the appearance of sameness before and after the wax melted was due to an intellectual perception that was not clear and distinct before. Only after reflective consideration does the meditator arrive at a certain judgment about his thinking. Mere consciousness of any given idea is not sufficient for knowledge of its content in cases where the idea we have is confused. Borrowing from the metaphor above, we see that not everything touched by the light of consciousness is necessarily known to its full extent.

But this does not exactly square with the initial assumption above that knowledge of our thoughts is instantly and utterly indubitable. It is only by having thoughts about thoughts, and clearly and distinctly bringing out what is in our minds – through introspection and reflection – that we attain knowledge of our mental affairs. Given this kind of evidence, it is difficult to preserve the strong interpretation of epistemic incorrigibility suggested by a cursory reading of the Second Meditation. While consciousness itself may be a necessary condition for ascribing knowledge to oneself, it is by no means sufficient. What is lacking is the connection between mere *consciousness* of our thoughts and *knowledge* of them, and this connection can be established only by means of reflective thinking. Thus, a transparent mind does not by itself entail incorrigibility for Descartes. We may presume that he adhered to some restricted form of epistemic privilege, the extent of which is perhaps closely tied to our preferred interpretation of his general strategy in the *Meditations*,³³ but whatever form that privilege takes, the entailment is by no means as clear as the unrefined interpretation of the transparency thesis would suggest. Keeping in line with these remarks, it would make sense to posit that the paradigmatic thoughts required in the Second Meditation to meet the meditator’s further epistemological goals are sufficiently clear and distinct to be beyond doubt.

As opposed to Descartes, we may safely posit that for Locke, consciousness of our mental states does entail knowledge. In order for us to do so, however, it needs to be shown that knowledge of our mental goings-on is due directly to conscious perception, as opposed to some further introspective act of attending to our thoughts. Let us briefly consider the evidence. According to Locke’s definition, knowledge is “nothing but *the perception of the connexion and agreement, or disagreement and repugnancy of any of our Ideas*. In this alone it consists.”³⁴ Of the several sorts of such agreements to be perceived, only the first – “*Identity, or Diversity*”³⁵ – is important for our present purposes. When we perceive any idea in the mind, we perceive it “to agree with it self, and to be what it is,”³⁶ and we therefore know that any idea in our minds is such as we perceive it to be. This pure, unmediated kind of knowledge is called *intuitive knowledge*.³⁷ Despite Locke’s somewhat unwieldy phrasing, we can draw from this not just the trivial claim that we perceive (and therefore know) every idea to be identical with itself, but also that we perceive (and therefore know) every idea precisely *as the idea* it is before the mind. He then repeatedly affirms that we require no further act of introspection in order to attain the highest degree of certainty when it comes to our own ideas: “clearly and infallibly” discovering their agreement is “the first Act of the

Mind” and it does this “without pains, labour, or deduction; but at first view, by its natural power of Perception and Distinction.”³⁸ We can therefore ascribe to Locke the position that mere perception of our ideas yields certain knowledge of them, because for him ideas as perceived by the mind are always distinct.³⁹ Moreover, it has already been established that consciousness is due to an internal, reflexive component in the perception of ideas: in perceiving any idea, we are conscious of it. And since it follows from the transparency thesis that we are conscious of all our thoughts, the result is that we have distinct epistemic privilege concerning the ideas in our minds.⁴⁰ In this respect, Locke was perhaps more Cartesian than Descartes himself.

But this conclusion should only increase our sense of unease. For if the thesis that all thinking is necessarily conscious strikes us as counterintuitive, then surely the thesis that we have knowledge of all our thinking seems even less credible. Let us therefore now turn to a qualifying aspect of mental transparency, namely, that it can come in different degrees. It seems that most of our skeptical intuitions concerning the transparency thesis stem from all those cases of unconscious or subconscious thinking or perceiving that we seem simply not to notice. Examples abound – for example, that annoying itch on my arm that I suddenly cease to be aware of when I bang my toe on the refrigerator. But the issue was pressed even in early modern times. In the Preface to the *New Essays*, in one of many attempts to dismantle the transparency thesis, Leibniz employs what Jolley calls the “argument from attention”:⁴¹

[W]hen we are not alerted, so to speak, to pay heed to certain of our own present perceptions, we allow them to slip by unconsidered and even unnoticed. But if someone alerts us to them straight away, and makes us take note, for instance, of some noise which we have just heard, then we remember it and are aware of just having had some sense of it. Thus, we were not straight away aware of these perceptions, and we became aware of them only because we were alerted to them after an interval, however brief.⁴²

The scenario described also seems intuitively plausible. Consider, for example, a thought-provoking conversation at dinner with the person seated next to you, during which you are too preoccupied to notice the noise of the restaurant, the clinking of cutlery, and the exchanges of the other guests at the table. Yet if you were asked immediately afterwards whether you could remember any of those things happening around you, it is probably the case that you could – or so Leibniz would argue. Transparency dictates that if those things were actually perceived by the mind (which they would have to be, considering the fact that they could be remembered later), they would have had to be conscious. But since they seem not to be, the transparency thesis must be false.

It would be mistaken, however, to think that Descartes and Locke, both obviously quite skilled at introspection, were oblivious to all those slight perceptions that one would not consider conscious either in Leibniz’s sense of the word⁴³ or in its everyday sense. In the case of Descartes, we should first clarify that the argument

from attention is geared against *occurrent* thoughts. Since Descartes is an innatist, he is already committed to the existence of inborn ideas of certain mathematical truths, God's existence, and so forth, the contents of which we are not always aware of, and even if we are aware of them we are not always certain of them. Descartes in all likelihood considered them to be unactualized dispositions and therefore not immediately (or in some cases at all) evident to the transparent mind. But even in the case of *occurrent* thoughts, Descartes is not defeated by Leibniz's argument: Simmons, for example, argues that there are degrees within what she calls "brute consciousness" (as opposed to "reflective consciousness"),⁴⁴ which correspond to the clarity of the conscious thoughts in question. "I call a perception clear," Descartes writes, "when it is present and accessible to the attentive mind."⁴⁵ Consequently, thoughts or perceptions that are less clear are also less present and accessible to the attentive mind, and it would stand to reason that when the mind is inattentive – as Leibniz's argument would demand – even more of our thoughts are only barely present and accessible. But this in no way precludes their being conscious in the way the transparency thesis dictates – they can later be remembered, after all! While he cannot agree with Leibniz in positing that these perceptions are unconscious, the move for Descartes would simply be to say that they were obscure, and were conscious in only a bare and rudimentary way, i.e. to a lower degree.

A similar view can also be ascribed to Locke, the difference being that little interpretative work is needed, since he is quite explicit on the issue. In discussing the different states of mind when thinking, Locke mentions that "the mind employs it self about them [i.e. ideas] with several degrees of Attention."⁴⁶ Sometimes, he continues, "it barely observes the train of *Ideas*, that succeed in the Understanding, without directing, and pursuing any of them: And at other times, it lets them pass almost quite unregarded, as faint shadows, that make no impression."⁴⁷ In these passages, Locke clearly qualifies the transparency thesis so as not to imply that all thinking is *equally* conscious. Naturally, the ideas perceived in my mind when I accidentally touch a hot stove are far more present to me than those perceived from the feeling on my skin of the clothes I am wearing. The latter ideas are so low on the perceptual spectrum, as it were, that calling them conscious strikes us as counterintuitive. But while these thoughts are indeed "dim and obscure to that degree, that they are very little removed from none at all,"⁴⁸ they are nonetheless conscious.

As is evident from the passages examined above, neither Locke's nor Descartes's version of the transparency thesis is needlessly strong. Their respective theories of mind can accommodate phenomena such as those raised in Leibniz's argument from attention or in our examples above. But the question remains: Why would they adhere to a thesis that, even when qualified in this way, some may find both highly problematic⁴⁹ and non-obvious? We should therefore turn to investigating the possible reasons for their shared commitment to transparency.

In the case of Descartes, his metaphysical considerations on the mind and body are a good start. For him, thinking is the sole and defining attribute of mind: the mind is *essentially* a thinking thing. Without thought, there is no mind. Moreover, thinking itself is defined in terms of consciousness: "I use [the term 'thought'] to include everything that is within us in such a way that we are immediately aware of

it.”⁵⁰ Considered this way, the answer to the question why Descartes subscribed to mental transparency is trivial: the mind is transparent by definition.

But perhaps there is a more satisfying answer to be found if we consider the question against the background of Descartes’s epistemological ambitions in the *Meditations*. The gist is the following. The meditator begins by questioning any of his beliefs for which he finds reason to doubt it. He then arrives at the “I think,” and from this new starting point begins to expand the realm of his knowledge to all his other thoughts, to God, to essences, and finally to objective reality. An important premise in this line of thought is that the mind is better known than the body, and perhaps even safe from doubt.⁵¹ Establishing this latter point is the goal of the Second Meditation. In this vein, it ends with the meditator concluding: “I know plainly that I can achieve an easier and more evident perception of my own mind than of anything else.”⁵² Any interpretation of the role of mental transparency must pay respect to these epistemological considerations on the self, which, as mentioned above, are not quite as clear as one may first be led to think. Mental transparency and epistemic privilege come apart. Nevertheless, it is certain that consciousness plays an important role in our knowledge of our own thoughts. Regardless of how exactly we spell out the degree to which Descartes was actually convinced that we have epistemic privilege concerning our own minds, it must be noted that mental transparency provides the bare principal condition of foundational knowledge. Consciousness extends as far as the mind reaches, and what is enclosed within consciousness defines the realm of what, for Descartes, any inquiry into the nature of reality must start with. Regardless of whether his primary intention was to defend against the skeptic or to draw a clear line between the mental realm and the corporeal,⁵³ consciousness provides a firm criterion for that of which foundational knowledge is even possible. The certainty of self-ascriptions, where permissible, is derived from careful reflective consideration against a backdrop of conscious accessibility.

But what of Locke? When he wrote his *Essay*, the intellectual world had changed considerably in the aftermath of Descartes’s *Meditations*. Under Cartesian influence, mental transparency had found its way into the works of thinkers such as La Forge and Malebranche. The question now is: What sort of advantage did Locke hope to gain by incorporating such a powerful premise into his philosophy? Quite a significant one, in fact: while Locke never argues explicitly for transparency, he gets a lot of mileage from its adoption by turning it against precisely the philosophical tradition from which it originated – that is, Cartesianism – most notably in his arguments against innatism and the thesis that the mind always thinks, i.e. that thinking is the essence of the soul. Moreover, statements of transparency figure prominently in Locke’s account of personal identity, thereby advancing a theory of personal identity that runs counter to the substantial metaphysics of his time. For these reasons, we should not be led astray by our everyday intuitions about the unconscious (or subconscious) and simply assume that transparency is a cumbersome relic that somehow slipped into Locke’s *Essay* due to his Cartesian roots, for this would seriously understate the role that it plays in his anti-Cartesian polemic.

4. Locke's Arguments⁵⁴

In this section, I will closely examine Locke's arguments against the cherished Cartesian doctrines of innate ideas and an ever-thinking soul. The role that mental transparency plays in this polemic becomes evident once we take into consideration the fact that (in the first edition of the *Essay* at least) mentions of the thesis pertain mainly – if not exclusively – to these arguments.⁵⁵ I will first examine the argument against innatism. In order to do so, it will be necessary to argue that memories – which might initially seem to be what is at stake with Locke's transparent mind – do not pose a problem for his version of the thesis. I will then turn to two separate but related arguments against the ever-thinking Cartesian soul.

4.1. Against Innate Ideas

Tracing Locke's polemic against innatism in its entirety would be a long and tedious task,⁵⁶ so I will concern myself only with the argument in chapter 4 of Book I of the *Essay*, where he proclaims that there is no sense in which an idea can be “in the mind” without being either remembered or occurrently perceived. He challenges the innatist by confronting her with a dilemma: “Whatever *Idea* is in the mind, is either an actual perception, or else having been an actual perception, is so in the mind, that by the memory it can be made an actual perception again.”⁵⁷ He then states that memories, when occurrently remembered, differ from regular perceptions in two ways: first, they must have been perceived previously, and secondly, this fact must be epistemically and phenomenally accessible to the subject that is remembering them. The latter difference is due to accompaniment by “a consciousness that it had been there before, and was not wholly a Stranger to the mind.”⁵⁸ In other words, an idea in the mind either is occurrently perceived and therefore – because of transparency – occurrently conscious, or was perceived at some previous point. This bifurcation is intended to contest the sense of being “in the mind” that the innatist requires to determine the ontological status of inborn ideas (e.g., as unactualized dispositions).

But this leads naturally to the question: How did Locke himself conceive of memory? For while it is obvious that mental transparency is the central assumption underlying the dilemma, it might be hard to see how this does not spell trouble for Locke himself. We should first get clear on why memories – which are also “in the mind” but not actually perceived – are not a problem for Locke's transparent mind. To do so, I will take my cue directly from Leibniz, who pointed out this apparent tension in the Preface to the *New Essays*:

[W]e are not always aware of our acquired dispositions or of the contents of our memory, and they do not even come to our aid whenever we need them, though often they come readily to mind when some idle circumstance reminds us of them.⁵⁹

At first glance, it would seem that a certain memory – say, of what I had for breakfast this morning – somehow involves various ideas (e.g., those of coffee and bread). Memories therefore seem to require the presence of ideas in the mind. But since we are not consciously aware of all our memories at any given time, it would

follow that treating memories as full-blown ideas in the mind is incompatible with the transparency thesis.

But it would be premature to accuse Locke of simply ignoring this obvious problem. As the following passage reveals, Locke acknowledges the fact that we are not constantly conscious of all our memories:

*Memory [...] is as it were the Store-house of our Ideas. For the narrow Mind of Man, not being capable of having many Ideas under View and Consideration at once, it was necessary to have a Repository, to lay up those Ideas, which at another time it might have use of.*⁶⁰

At first glance, this solution is less than elegant. For Locke simply denies memories the status of mental items, and even seems to deny them any ontological status at all when they are not actively remembered or brought into view:

But our *Ideas* being nothing, but actual Perceptions in the Mind, which cease to be any thing, when there is no perception of them, this *laying up* of our *Ideas* in the Repository of the Memory signifies no more but this, that the Mind has a Power, in many cases, to revive Perceptions, which it has once had [...]. And in this Sense it is, that our *Ideas* are said to be in our Memories, when indeed, they are actually no where, but only there is an ability of the Mind, when it will, to revive them again.⁶¹

It is clear from historical context that Locke added this passage in the second edition of the *Essay* in order to defend his conception of memory against a critique made by John Norris, who argued against it on similar grounds.⁶² The language used here is unambiguous: memories “cease to be anything” and are “actually nowhere” when not remembered. What Locke is claiming is that the mind has a certain capacity to remember, and it is only when that faculty is exercised that these past ideas are again perceived, i.e. are again conscious.

But one might wonder how this faculty is supposed to operate upon quite literally nothing. Even clarifying the account by positing that memories are dispositions seems not to help: memories are commonly considered thoughts in the proper sense with a phenomenal or propositional content, and simply rephrasing this to mean “dispositions to have occurrent thoughts with a phenomenal or propositional content” does not explain what those dispositions are grounded in. Rather, it merely shifts the problem, such that the relevant question becomes: Where are the contents of our memories stored (to use Locke’s own metaphor)? In the corresponding part of the *New Essays*, Leibniz expresses similar doubts about Locke’s proposed solution:

What is needed is a somewhat clearer explanation of what this faculty consists in and how it is exercised [...]. If nothing were left of past thoughts the moment we ceased to think of them, it would be

impossible to explain how we could keep the memory of them; to resort to a bare faculty to do the work is to talk unintelligibly.⁶³

Before searching for an answer in Locke's own text, it is perhaps helpful first to turn to Descartes, who faced the same problem but was much more explicit in proposing a solution. Notably, Descartes did not have a singular concept of memory, but rather distinguished between corporeal and intellectual memory, the latter being either memory of intellectual (i.e. non-sensory) ideas, or memory stored in the intellect. In a letter to Mesland from May 2, 1644, he writes:

As for memory, I think that the memory of material things depends on the traces which remain in the brain after an image has been imprinted on it; and that the memory of intellectual things depends on some other traces which remain in the mind itself. [...] The traces in the brain [...] dispose it to move the soul in the same way as it moved it before, and thus to make it remember something.⁶⁴

Indeed, it is just this distinction that poses its own problems for Cartesian transparency, simply because it is explicitly stated that it does not admit of the clean solution provided for the case of sensory memory immediately beforehand – namely, that sensory memories are grounded in traces in the brain.⁶⁵ More importantly, however, this further complication is not relevant to the Lockean account of memory, for Locke does not draw any such distinction.

With a possible solution to the problem whether the transparency thesis can be reconciled with the storehouse metaphor of memory, the next question would be: Is there any evidence for it in the *Essay*? The answer is yes. Weinberg argues that there is, in fact, textual evidence suggesting that Lockean memories, when not remembered as conscious, occurrent thoughts, are stored as traces in the brain, which can later be revived by an effort of the mind.⁶⁶ For example, in describing how songbirds retain melodies sung previously, Locke notes:

Since there is no reason why the sound of a Pipe should leave traces in their brains, which not at first, but by their after-endeavours, should produce the like Sounds; and why the Sounds they make themselves, should not make traces which they should follow, as well as those of the Pipe, is impossible to conceive.⁶⁷

On the assumption that the memory of songbirds is similar to that of more developed animals, it seems that impressions are stored as traces in the brain, to be revived by our “after-endeavours,” i.e. the workings of our faculty of memory. These and similar passages suggest that Locke could save himself from Leibniz's argument by positing that memories are stored physically, as opposed to mentally. We would then not be constantly conscious of them, because they are not mental items, but rather physical ones. Weinberg's solution moreover fits quite neatly with Descartes's treatment of corporeal memory.⁶⁸

With this qualification, then, the overall argument provides Locke with a strong case against the Cartesian agenda. Providing a coherent account of how memories can be said to be “in the mind” when not occurrently perceived allows Locke to challenge the innatist with the dilemma mentioned above, by closing off the possibility of there being anything in the mind that is neither perceived now nor was perceived previously. It is mental transparency that allows Locke to present the innatist with a forced choice, and his account of memory that prevents him from falling victim to his own dilemma. This argument strengthens his anti-innatist polemic, since it is also applicable (albeit questionably) to theories in which innate ideas are unactualized dispositions, which the rest of his arguments in Book I of the *Essay* would miss.

4.2. Against an Ever-Thinking Soul

Let us now turn to the second anti-Cartesian use to which Locke puts the transparency thesis: his polemic against the thesis that the mind always thinks. His motivation for rejecting the thesis is twofold. First, he argues against it because it is untenable according to his own views on substance. As already mentioned, Descartes is convinced that the soul always thinks, because the soul is essentially a thinking thing. But since on Locke’s account the nature of substance is unknown to us, to the extent that it is even epistemically possible for matter to think, the thesis of an ever-thinking soul is too bold a metaphysical claim. Locke’s own considerations on the issue reveal that for him the question must be settled empirically: “But whether that Substance perpetually thinks, or no, we can be no farther assured, than Experience informs us.”⁶⁹ Second, the doctrine entails a hidden argument for innatism, which Locke hoped to have refuted in Book I of the *Essay*. For if the soul *always* thinks, it would follow that it also entertains thoughts from the very moment of its creation – that is, before it could have acquired any ideas from sensation or reflection. And if there were any ideas in the mind that did not stem from either sensation or reflection, Locke’s concept empiricism would be compromised, since the innatist could point to them as innate.

We can distinguish a variety of different arguments Locke employs to refute the thesis that the mind always thinks, but the two that are perhaps the most cogent both make use of the transparency thesis. Let us turn first to the argument from dreamless sleep. It appeals to the fact that we seem not to be thinking while we are asleep and not dreaming: “’Tis doubted whether I thought all last night, or no; the Question being about a matter of fact, ‘tis begging it to bring, as a proof for it, an Hypothesis which is the very thing in dispute.”⁷⁰ Locke thus considers it fallacious to conclude that the soul thinks during dreamless sleep simply on the basis of the supposition that it thinks essentially, and the thesis itself is evidently proven false by this example. But to give a positive answer as to how we can know that the soul does not think during dreamless sleep, Locke must allude to the transparency thesis, arguing that it is impossible for us to think or to perceive without being conscious of it: We “cannot think at any time waking or sleeping, without being sensible of it.” With mental transparency as a premise, Locke has merely to call on the reader to introspect and to discover that we seem not to think while asleep, thus allowing Locke

to expose the opposing view as question-begging. But he can insist on settling the matter empirically only by asserting that the connection between consciousness and thought is necessary. For if he conceded that the soul can think without being conscious of it, a proponent of the thesis could simply press this point and say that the mind does in fact do so, at least in sleep.

Interestingly, in chapter 1 of Book II of the *Essay* Locke even considers a response to his claim based on an argument made by Descartes in his Replies to the Fourth Set of Objections by Arnauld.⁷¹ It might be objected against Locke that the soul does in fact think during dreamless sleep but that those thoughts are immediately forgotten. Even though this counterargument is notorious for its ad hoc character, Locke takes it at face value.⁷² His reasoning against this objection is that having such thoughts would be excessively uneconomical: “*To think often, and never to retain it so much as one moment, is a very useless sort of thinking.*”⁷³ Locke claims that these thoughts would serve no functional purpose, because we cannot recall or interact with them in any way that is helpful to conducting our daily lives. His point is reinforced by an appeal to theological pragmatism: Why would God create man with a faculty of thinking that is “so idly and uselessly employ’d” for so many hours every day? The claim that thought should have some useful purpose is especially important, since it also implicitly makes it seem much less credible that thoughts entertained during dreamless sleep could be conscious, even if only to a low degree (see above).

In the context of Locke’s polemic against an ever-thinking soul, the transparency thesis allows him to dismantle Cartesian substantial metaphysics by insisting on settling the matter by simple observation and substituting the bold metaphysical claims of Descartes with his own much more modest account, according to which thinking is an activity rather than the essence of the soul. Leibniz would later, in an effort to vindicate the originally Cartesian doctrines of innatism and the ever-thinking soul, devote a considerable amount of space in the *New Essays* to his own polemic against Locke’s views. Leibniz’s main goal in the Preface and Books I and II is to advance his theory of *petites perceptions* – minute perceptions below the threshold of consciousness – thereby combating the notion of mental transparency in order to pry open the door once again for a full-fledged rationalist and substantial metaphysics of mind.⁷⁴ This further illustrates the importance of mental transparency as a tool for Locke to claim and safeguard his empiricist principles against his rationalist adversaries.

4.3. Personal Identity

In this last section on Locke’s arguments against Cartesian metaphysics, I will turn to the transparency thesis as it figures in personal identity. We can identify two separate prongs of Locke’s attack: first, the above-mentioned use of personal identity in a concrete argument against the ever-thinking Cartesian soul, and secondly, the broader role of Locke’s consciousness-based account of personal identity as an alternative to a substance-based theory. This latter point, however, will have to be considered only briefly, since personal identity – and more specifically the role consciousness plays in it – are independent issues in the scholarly literature,⁷⁵ and subject to their own intricacies.⁷⁶

Concerning the former, Locke's intention is to drive home the point that the Cartesian views on the soul are incompatible with transparency. The argument is similar in spirit to the one examined in the preceding section, in that it advances the claim that Descartes's position (this time if considered by the lights of what it entails for personal identity) opens up a host of dubious possibilities. It comes in the form of a thought experiment based on three Cartesian premises, the first two of which we are already familiar with – namely, that the soul is essentially thinking and that all thought is conscious; the third, Locke posits, is that on the Cartesian account it should be metaphysically possible for bodies to exist without their souls, and vice versa. In the *Essay*, he draws this conclusion from Descartes's well-known position that animals have no souls: This “is no impossible Supposition for the Men [...] who so liberally allow Life, without a thinking Soul to all other Animals.”⁷⁷ Indeed, Descartes locates the source of selfhood in a person's immaterial, ever-thinking soul: “I can infer correctly that my essence consists solely in the fact that I am a thinking thing. It is true that I may have [...] a body that is very closely joined to me.”⁷⁸ This rather weak link of “being closely joined” points to the body not being a constitutive element of selfhood, properly speaking. In the *Discourse*, Descartes further reinforces the possibility raised by Locke, writing that “the soul by which I am what I am – is entirely distinct from the body, [...] and would not fail to be whatever it is, even if the body did not exist.”⁷⁹

Locke then puts forward the following challenge: since the Cartesian position is firmly committed to the theses that the soul always thinks, that this thinking is – *qua* transparency – always conscious to its subject, and that it is metaphysically possible for the soul to think apart from its body, it should follow that our souls do not inhere in our bodies while we are asleep. Moreover, this implies the further possibility of our souls inhabiting other bodies and doing their thinking there: “For if *Castor's* Soul can think whilst *Castor* is asleep, what *Castor* is never conscious of, 'tis no matter what Place it chuses to think in.”⁸⁰ Here Locke implicitly appeals to the assumption that our own conscious experience would necessarily inform us of any thinking that goes on while we are asleep. Since (for obvious reasons) it does not so inform us, our soul would have to think apart from us. His point in this first half of the polemic is that Descartes would presumably prefer to avoid the possibility of such scenarios.

But Locke then asks where we are to place personal identity in cases where two persons share one soul:

I ask then, Whether *Castor* and *Pollux*, thus, with only one Soul between them, which thinks and perceives in one, what the other is never conscious of, nor is concerned for, are not two as distinct Persons, as *Castor* and *Hercules*; or as *Socrates* and *Plato* were?⁸¹

On the one hand, Locke seems simply to be pressing his preceding point, adding to the already numerous counterintuitive consequences of Cartesian metaphysics; on the other hand, however, he subtly points to the inadequacies of substance as a general criterion for personal identity. Although personal identity may not be a traditional point of contention between Locke and Descartes, it becomes

obvious over the course of Locke's polemic that even more cracks emerge in the Cartesian picture. Descartes and his followers would have simply pointed to the soul as that which guarantees personal identity,⁸² but Locke discovered that there are questions, such as those above, that are left unanswered by this view. These questions must have intrigued Locke himself, since he elaborated upon his claims in the now famous chapter 17 of Book II of the *Essay*, which he added in its second edition in 1694. This, then, is the second prong of Locke's polemic and the positive thesis contained therein – that is, the departure from immaterial substance as an objective means of determining the identity of persons over time, and the subsequent shift to consciousness as a criterion that is accessible from a first-person, subjective point of view.

The upshot of Locke's initial arguments presented in this subsection is an innovative attempt to dismantle the Cartesian doctrine of the ever-thinking soul. But even more important is the crucial role they subsequently play in paving the way for an entirely original theory of what constitutes the self, which is ultimately motivated by Locke's conviction that the metaphysics of personhood should be moved back into the realm of our experience.

5. Conclusion

In the introduction to this paper, I presented the problems associated with the thesis of mental transparency by alluding to our uneasy intuitions concerning the seemingly obvious counterexamples our experience provides. One could easily have been led to think that the thesis is an odd concept, one that is philosophically outdated and empirically on shaky ground. But this impression is misguided, for it fails to take into account what mental transparency accomplished in the philosophical context of its time. I hope to have shown how the premise of mental transparency – which Locke shares with Descartes in the context of their respective theories of consciousness, which are remarkably similar in structural terms – is used by Locke to great effect against his adversary. The *Essay*'s overarching theme of turning to introspection and experience rather than bold metaphysical claims – such as those about innate ideas and essentially thinking souls – to explain the working of the human mind is underscored by cogent arguments aimed at exposing the Cartesian view as either question-begging or plagued by all sorts of odd consequences. And although mental transparency does not square well with our post-Freudian notions, its historical impact can hardly be overestimated.

Acknowledgments. I would like to thank Martin Klein, Naomi Osorio-Kupferblum, and Olivér István Tóth for organizing the Central Kolleg workshop on 'Consciousness in Medieval and Early Modern Philosophy' in December 2017, Berlin, where I first presented this paper, and for their helpful comments and insights. I am also grateful to the journal's two anonymous reviewers for their feedback, which greatly improved my paper.

References

- ¹ All citations to Descartes take the form “AT volume:page/CSM volume:page,” referring to *Oeuvres de Descartes*, ed. C. Adam and P. Tannery, 11 vols. (Paris: Vrin, 1996), as translated in *The Philosophical Writings of Descartes*, trans. J. Cottingham, R. Stoothoff, D. Murdoch, and A. Kenny, 3 vols. (Cambridge: Cambridge University Press, 1985–1991). For Locke’s *Essay*, citations take the form “ECHU book.chapter.section” and refer to Locke, J., *An Essay Concerning Human Understanding*, ed. P. H. Niddich. (Oxford: Oxford University Press, 1975). References to Leibniz are given in the form “NE book.chapter.section,” referring to Leibniz, G., *New Essays on Human Understanding*, trans. P. Remnant and J. Bennett (Cambridge: Cambridge University Press, 1996). Note that this translation reproduces the pagination from the Akademie-Verlag edition (abbreviated as “A VI.vi”): Leibniz, G., *Nouveaux essais sur l’entendement humain*, ed. A. Robinet and H. Schepers, *Sämtliche Schriften und Briefe, Reihe VI (Philosophische Schriften)*, vol. vi (Berlin: Akademie-Verlag, 1962).
- ² The term “transparency” is sometimes used, especially in the literature on Descartes, to refer specifically to *epistemological* transparency, i.e. the doctrine that we have knowledge of all our thinking. In this paper I prefer to use “incorrigibility” or “epistemic privilege” when referring to this closely related thesis, and I restrict my use of “transparency” to consciousness. I distinguish these two concepts in section 2.
- ³ Simmons, A., “Cartesian Consciousness Reconsidered”, *Philosopher’s Imprint* 12/2 (2012): 1–21, at 2.
- ⁴ ECHU, Introduction, 8.
- ⁵ AT VII:366/CSM II:253.
- ⁶ Because of the overwhelming volume of literature on ideas in early modern philosophy in general, as well as in Locke and Descartes specifically, I restrict myself to the most basic statements about the relationship of ideas to consciousness. For an overview, see Ayers, M., “Ideas and Objective Being”, in *The Cambridge History of Seventeenth-Century Philosophy*, ed. D. Garber and M. Ayers (Cambridge: Cambridge University Press, 1998), 1062–1107; see esp. n2 for further reading on the topic. For an insightful analysis of Lockean ideas, see Bolton, M., “The Taxonomy of Ideas in Locke’s Essay”, in *The Cambridge Companion to Locke’s Essay*, ed. L. Newman (Cambridge: Cambridge University Press, 2007), 67–100.
- ⁷ For an in-depth historical analysis of these themes, see Weinberg, S., *Consciousness in Locke* (Oxford: Oxford University Press, 2015), ch. 1.
- ⁸ Udo Thiel links the Lockean views on consciousness and reflection closely to La Forge and Arnauld, but he does not share the contention presented here that Descartes was himself a first-order theorist of consciousness. See Thiel, U., *The Early Modern Subject: Self-Consciousness and Personal Identity from Descartes to Hume* (Oxford: Oxford University Press, 2011), 35–60, 117. For a more detailed discussion of first- and second-order theories of consciousness in Descartes, see Lähteenmäki, V., “Orders of Consciousness and Forms of Reflexivity in Descartes”, in *Consciousness: From Perception to Reflection in the History of Philosophy*, ed. S. Heinämaa, V. Lähteenmäki, and P. Remes (Dordrecht: Springer, 2007), 177–201 and Simmons, A., (2012).
- ⁹ AT VII:160/CSM II:130 (my italics).
- ¹⁰ AT VII:181/CSM II:137 (my italics).
- ¹¹ AT XI:343/CSM I:335–336.
- ¹² An example of a piece of textual evidence for the contrary claim that Descartes was a higher-order theorist is the *Conversation with Burman*, where he says: “It is correct that to be aware is both to think and to reflect on one’s thought. But it is false that this reflection cannot occur while the previous thought is still there. This is because, as we have already seen, the soul is capable of thinking of more than one thing at the same time, and of continuing with a

particular thought which it has. It has the power to reflect on its thoughts as often as it likes, and to be aware of its thought in this way” (AT V:149/CSM III:335). Cf. note 8 above.

¹³ La Forge, L. de, *Treatise on the Human Mind*, trans. D. M. Clarke (Dordrecht: Kluwer, 1997), 57.

¹⁴ Lähteenmäki, V., (2007).

¹⁵ Arnauld, A., *On True and False Ideas*, trans. S. Gaukroger (Manchester: Manchester University Press, 1990), 71.

¹⁶ For more in-depth considerations on the historical relationship between Descartes’s and Arnauld’s theories of consciousness, see Daniel Schmal’s paper in this issue.

¹⁷ See Coventry, A. and Kriegel, U., “Locke on Consciousness”, *History of Philosophy Quarterly* 25/3 (2008): 221–242; Lähteenmäki, V., “The Sphere of Experience in Locke: The Relations between Reflection, Consciousness, and Ideas”, *Locke Studies* 8/1 (2008): 59–100; Thiel, U., “Leibniz and the Concept of Apperception”, *Archiv für Geschichte der Philosophie* 76/2 (1994): 195–209; see also Thiel, U., (2011); Weinberg, S., “The Coherence of Consciousness in Locke’s Essay”, *History of Philosophy Quarterly* 25/1 (2008): 21–39; and Weinberg, S., (2015). For the higher-order reading, see Kulstad, M., *Leibniz on Apperception, Consciousness, and Reflection* (Munich: Philosophia, 1991) and McRae, R., *Leibniz: Perception, Apperception, and Thought* (Toronto: University of Toronto Press, 1976).

¹⁸ ECHU II.i.19.

¹⁹ For this argument, see Coventry, A. and Kriegel, U., (2008): 232. This contingency is best spelled out in terms of the possibility of the numerically distinct mental states, that is, the lower-order thought and the higher-order thought, existing independently of one another. If the lower-order thought exists without a higher-order thought, then it is simply unconscious; if the higher-order thought exists without a corresponding lower-order thought, it will appear to the subject, paradoxically, that she is having conscious experience without actually having it. For discussion, see Lycan, W., “Consciousness as Internal Monitoring”, in *The Nature of Consciousness: Philosophical Debates*, ed. N. Block, O. Flanagan, and G. Güzeldere (Cambridge, MA: MIT Press, 1997), 755–771, esp. 756–757.

²⁰ ECHU II.xxvii.9 (my italics).

²¹ NE II.i.19, A VI.vi:118.

²² It should be noted, however, that some early modern philosophers actually embraced an infinite regress of this sort. A good example is Spinoza, who explicitly mentions this infinite regress of thoughts having other thoughts as their object (see the scholium to proposition 21 in part two of the *Ethics* in *The Collected Works of Spinoza*, 2 vols., ed. and trans. E. Curley, (Princeton: Princeton University Press, 1988–2016)). According to the interpretation proposed by Thiel, Descartes held a similar view; see Thiel, U., (2011): 35–60, 117, and cf. notes 8 and 12 above.

²³ AT VII:246/CSM II:171.

²⁴ ECHU II.i.12.

²⁵ This is sometimes referred to as “epistemic transparency.” See note 2 above.

²⁶ AT VII:29/CSM II:19.

²⁷ ECHU II.xxvii.9.

²⁸ AT VII:28/CSM II:19; cf. AT VII:29/CSM II:19.

²⁹ The debate on Descartes’s conception of self-knowledge and its role in the *Meditations* is complex and still unresolved. For influential discussions, see McRae, R., “Descartes’ Definition of Thought”, in *Cartesian Studies*, ed. R. J. Butler (Oxford: Blackwell, 1972), 55–70; Curley, E., *Descartes Against the Skeptics* (Cambridge, MA: Harvard University Press, 1978), 170–193; Wilson, M., *Descartes* (New York: Routledge, 1978), 150–165; Radner, D., “Thought and Consciousness in Descartes”, *Journal of the History of Philosophy* 26/3 (1988): 439–452; Alanen,

L., *Descartes's Concept of Mind* (Cambridge, MA: Harvard University Press, 2003), 53–56, 99–102; Rozemond, M., “The Nature of Mind”, in *The Blackwell Guide to Descartes' Meditations*, ed. S. Gaukroger (Oxford: Blackwell, 2006), 48–66; Broughton, J., “Self-Knowledge”, in *A Companion to Descartes*, ed. J. Broughton and J. Carriero (Oxford: Blackwell, 2008), 179–195.

³⁰ AT VI:23/CSM I:122.

³¹ See Rozemond, M., (2006), 59–61. Rozemond's other examples include the textual context of the already-mentioned passage from the *Discourse on Method* (AT VI: 23; CSM I: 122), the *Passions of the Soul*, (AT XI:349–350/CSM I:338–339), and the *Principles of Philosophy* (AT VIIIA:35–36/CSM I:218–219).

³² AT VII:31/CSM II:21.

³³ Some commentators such as Curley have primarily emphasized the role incorrigibility plays in arguing against the skeptic, while others such as Wilson frame its main purpose as supporting Descartes's dualist agenda. See Curley, E. (1978), 35–45 and Wilson, M., (1978), 1–11; for a critical perspective on both, see Rozemond, M., (2006), 57–66.

³⁴ ECHU IV.i.2.

³⁵ ECHU IV.i.3.

³⁶ ECHU IV.i.3.

³⁷ ECHU IV.ii.1.

³⁸ ECHU IV.i.4. See also ECHU IV.ii.1, where Locke writes that “the Mind is at no pains of proving or examining, but perceives the Truth as the Eye doth light, only by being directed towards it. [...] Such kinds of Truths, the Mind perceives at the first sight of the *Ideas* together, by bare *Intuition*; without the intervention of any other *Idea*; and this kind of Knowledge is the clearest, and most certain, that human Frailty is capable of.”

³⁹ Strictly speaking, there are no “confused ideas” before the mind on Locke's account. He writes that an idea “can be no other but such as the Mind perceives it to be” (ECHU II.xxix.5). The only source of confusion that Locke allows for is when we incorrectly apply language to our ideas: “[T]hat which makes it *confused* is, when it is such, that it may as well be called by another Name, as that which it is expressed by, the difference which keeps the Things (to be ranked under those two different Names) distinct, and makes some of them belong rather to the one, and some of them to the other of those Names, being left out; and so the distinction, which was intended to be kept up by those different Names, is quite lost” (ECHU II.xxix.6).

⁴⁰ For an extensive argument that knowledge of individual ideas is due to consciously perceiving them and that this knowledge is propositional, see Weinberg, S., (2015), 61–73.

⁴¹ Jolley, N., *Leibniz* (London and New York: Routledge, 2005), 120.

⁴² NE Preface, A VI.vi:54.

⁴³ Consciousness in its modern usage is more similar to Leibniz's term “apperception,” which serves to distinguish consciousness proper from minute unconscious perceptions. His views on this matter are most clearly stated in the Preface and Book II of the *New Essays*.

⁴⁴ Simmons, A., (2012): 16–17; cf. Lähteenmäki, V., (2007) on the distinction between rudimentary and reflexive consciousness mentioned in section 1 above.

⁴⁵ AT VIIIA:22/CSM I:208.

⁴⁶ ECHU II.xix.3.

⁴⁷ ECHU II.xix.3.

⁴⁸ ECHU II.xix.4.

⁴⁹ See Simmons, A., (2012) for a comprehensive analysis of the issues that transparency raises for Descartes even from the perspective of his own philosophy.

⁵⁰ AT VII:160/CSM II:113, cf. AT VIIIA:7/CSM I:195.

⁵¹ This may be simply false, however. Wilson, for example, accuses Descartes of being “[far] from giving the Deceiver Hypothesis the full force that seems, logically, to be implicit in it” (Wilson, M., (1978), 152).

⁵² AT VII:34/CSM II:22–23.

⁵³ See notes 29 and 33 above.

⁵⁴ Weinberg, S., (2015) discusses several of these arguments over the course of her excellent work on Lockean consciousness. I am especially indebted to her insightful discussions of memory in subsection 1, 169–173, and her remarks on personal identity in subsection 3, esp. 145–147.

⁵⁵ They also feature prominently in *ECHU* II.xxvii, which was added in the second edition; see Nidditch’s historico-critical notes in Locke, J., (1975).

⁵⁶ For a detailed discussion of Locke’s polemic against innatism in its entirety, see Rickless, S., “Locke’s Polemic against Nativism”, in *The Cambridge Companion to Locke’s Essay*, ed. L. Newman (Cambridge: Cambridge University Press, 2007), 33–66. For an interpretation of the role of consciousness in the argument under discussion, see Weinberg, S., (2015), 47–50.

⁵⁷ *ECHU* I.iv.20.

⁵⁸ *ECHU* I.iv.20.

⁵⁹ *NE* Preface, A VI.vi:52.

⁶⁰ *ECHU* II.x.2.

⁶¹ *ECHU* II.x.2.

⁶² As James Gibson points out, this counterargument was first proposed in Norris’s *Cursory Reflections upon a Book Called an Essay Concerning Human Understanding*, published in 1690 as an appendix to his book *Christian Blessedness*. See Gibson, J., *Locke’s Theory of Knowledge and Its Historical Relations* (Cambridge: Cambridge University Press, 1917), 23.

⁶³ *NE* II.x.2, A VI.vi:140.

⁶⁴ AT IV:114/CSM III:233.

⁶⁵ Simmons argues that Descartes’s innate ideas and intellectual memory are not in the mind due to its faculty to conjure them up in accordance with certain dispositions. Her proposal is that the mind comes to have these ideas via higher-order contemplation of occurrently conscious ideas. See Simmons, A., (2012): 18–19.

⁶⁶ Weinberg, S., “The Metaphysical Fact of Consciousness in Locke’s Theory of Personal Identity”, *Journal of the History of Philosophy* 50/3 (2012): 387–415, at 410–412. See also Weinberg, S., (2015), 170–173.

⁶⁷ *ECHU* II.x.10.

⁶⁸ Moreover, Weinberg’s answer as to why Locke would dodge the issue, as it were, lends further credibility to this parallel: “[I]t is consistent with his agnosticism with respect to the nature of thinking substance and with respect to the metaphysics of causal relations not to delve into any substantive explanation” (Weinberg, S., (2015): 172). Descartes, who has no such qualms concerning bold claims about the nature of body and mind, can simply state them outright.

⁶⁹ *ECHU* II.i.10.

⁷⁰ *ECHU* II.i.10.

⁷¹ AT VII:214/CSM II:150, AT VII:246/CSM II:171–172.

⁷² Simmons to some extent defends this claim: “In the end, Descartes only appeals to the vagaries of memory to argue for the existence of fleeting thoughts in infants and sleeping adults that are unaccompanied by other thoughts that serve to make them salient, and so remembered, in our cognitive lives” (Simmons, A., (2012): 13–14).

⁷³ *ECHU* II.i.15.

⁷⁴ Leibniz's counterarguments appear in the corresponding sections of the *New Essays*; see *NE* Preface and II.i.11–19. For an insightful analysis of his polemic, see Jolley, N., *Leibniz and Locke: A Study of the New Essays on Human Understanding* (Oxford: Clarendon Press, 1984).

⁷⁵ Weinberg distinguishes two interrelated kinds of consciousness in Locke's *Essay*: on the one hand, the reflexive awareness inherent in every thought (see section 1 above), and on the other hand, the "metaphysical fact of consciousness" that figures in his theory of personal identity. See Weinberg, S., (2015), 153.

⁷⁶ For an excellent overview of both the historical and contemporary debate, see Thiel, U., (2011), 121–221.

⁷⁷ *ECHU* II.i.12.

⁷⁸ *AT* VII:78/*CSM* II:54.

⁷⁹ *AT* VI:33/*CSM* I:127.

⁸⁰ *ECHU* II.i.12.

⁸¹ *ECHU* II.i.12.

⁸² For a more detailed discussion of pre-Lockean theories of personal identity in the context of substantial metaphysics, see Thiel U., (2011), 35–43, 224–276.