

IS LEIBNIZ'S THEORY OF 'I' COHERENT?

MIN HEO 
UC San Diego

SAMUEL C. RICKLESS 
UC San Diego

As Margaret Wilson famously argued, Leibniz appears committed to two mutually inconsistent propositions: (1) I am a particular substance, and (2) It is metaphysically possible that I continue to exist independently of this particular substance. Three solutions have been offered to this conundrum on Leibniz's behalf: that (1) and (2) are compatible because I am only contingently identical with this particular substance; that Leibniz rejects (1); and that Leibniz rejects (2). In this essay, we explain why Leibniz does not accept any of these solutions, and we then offer two additional textually supported ways for Leibniz to avoid inconsistency. On the 'modal' solution, Leibniz identifies the fact that makes (2) true as a fact about my counterpart and a counterpart of the substance with which I am identical, and hence (1) and (2) can both be true together. On the 'ambiguity' solution, Leibniz thinks that 'I' can be used to refer to a soul or to a person, and he endorses (1) only when 'I' refers to a soul and endorses (2) only when 'I' refers to a person. On either the modal solution or the ambiguity solution, Leibniz avoids inconsistency and Wilson's conundrum is solved.

Keywords: Leibniz; personal identity; self; person; substance

Contact: Min Heo <miheo@ucsd.edu>

Journal of Modern Philosophy is a peer-reviewed open access journal published by the Aperiio. © 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



OPEN ACCESS

1. Is Leibniz's Theory of 'I' Incoherent?: Wilson's Challenge

Gottfried Wilhelm Leibniz wrote *Nouveaux Essais* (1704; henceforth, the *New Essays*) in response to John Locke's *An Essay concerning Human Understanding* (1689/1975; henceforth, *Essay*). One of the issues with which Leibniz was dissatisfied in the *New Essays* was Locke's discussion of personal identity. According to Locke, a person is defined as a rational being capable of self-reflection and of considering herself the same entity throughout time: 'What *Person* stands for, ... is a thinking intelligent Being, that has reason and reflection, and can consider it self as it self, the same thinking thing in different times and places' (E II.27.9). At the same time, Locke famously separates the issue of personal identity from that of the identity of substance or soul: 'For it being the same consciousness that makes a Man be himself to himself, *personal Identity* depends on that only, whether it be annexed only to one individual Substance, or can be continued in a succession of several Substances' (E II.27.10). Locke maintains that personal identity is based solely on the identity of consciousness, which is founded on the continuity of memory. For Locke, whether X at t_1 is the same person as Y at t_2 is purely a psychological matter (i.e., whether Y at t_2 remembers X 's experience at t_1 as their own) and is independent of the metaphysical issues centered on the notion of substance.

Leibniz, in contrast, maintains throughout his philosophical career that substance plays a crucial role in the theory of personal identity. For instance, in the draft of his July letter to Arnauld (1686), Leibniz holds that the identity of substance is a necessary condition for the sameness of a person through time:

Let us then first of all take me as existing during the time AB, and also as existing during the time BC. Since then one supposes that ... it is I who exist in the time AB and am then in Paris, and that it is also I who exist in the time BC and am then in Germany, *there must of necessity be a reason for the true statement that we continue to exist, that is to say that I who was in Paris am now in Germany*. For if there is no reason, one would be as justified in saying that it is another person. To be sure, my subjective experience has convinced me *a posteriori* of this identity, but there must also be one *a priori*. Now, it is impossible to find another identity, except that my attributes of the preceding time and state as well as those of the following time and state are predicates of one and the same subject, they are present in the same subject. (G II.43/LA: 46–47; our emphasis)

Here Leibniz emphasizes that although my subjective experiences (e.g., memories of the past) make me convinced of my diachronic identity, that identity

should also be grounded in *a priori* reason.¹ For Leibniz, the requisite reason cannot be found in anything other than that the distinct attributes I have at different times belong to the 'same subject'—i.e., the same substance. Similarly, Leibniz emphasizes in the *New Essays* that the order of nature makes personal identity always presuppose the identity of substance: 'You [i.e., Locke] seem to hold, sir, that this apparent identity could be preserved in the absence of any real identity. ... I should have thought that, according to the order of things, an identity which is apparent to the person concerned—one who senses himself to be the same—presupposes a real identity' (NE: 236).² This emphasis on the sameness of substance, notably, brings Leibniz closer to the Cartesians, who regard 'I' as a thinking substance, rather than to the Lockeans, who separate the issue of personal identity from that of the identity of substance.

At the same time, although Leibniz reveals his dissatisfaction with Locke's treatment of personal identity, his discussion of the immortality of the soul contains features favorable to Lockeans and ultimately locates Leibniz at a midpoint between Cartesians and Lockeans. Descartes maintains that the natural immortality of the soul follows from the indestructibility of substances (AT VII.13–14).³ But for Leibniz, the indestructibility of substances is not enough for any meaningful immortality. The immortality of the soul is typically postulated to make divine justice compatible with the fact that there are persons who are not rewarded in their lifetime for their goodness or who are not punished for their wickedness. However, divine punishment or reward would not be fully justifiable if moral agents did not retain memories of their merits and demerits and, consequently, could not see those merits and demerits as their own. Thus, Leibniz holds that immortality additionally requires preserving the sameness of personality (T: 89).⁴ That is, Leibniz agrees with Locke that what matters from the moral and religious points of view is personal identity, not sameness of substance. Accordingly, Leibniz takes a 'hybrid' view concerning 'I' in that the identity of substance is essential in addressing metaphysical issues such as the subsistence of 'I', while personal identity plays an indispensable role

1. It is worth noting that Leibniz often uses the term '*a priori*' in its older sense. According to Robert Adams, '[p]roofs *a priori* and *a posteriori*, in the original sense of those terms, are proofs from the cause and from the effects, respectively, of the fact to be proved' (Adams 1994: 109). In the *New Essays*, Leibniz uses the term the way Adams describes it: Reason 'would make known the reality [of a definition] *a priori* in exhibiting the cause or possible generation of the thing defined' (NE: 294).

2. Instead of the term 'personal identity,' Leibniz often uses 'apparent identity'. Likewise, 'real identity' and 'physical identity' are Leibniz's terminology to denote the identity of substance. Concerning Leibniz's explanation of these terms and how they relate to the 'substantial identity/personal identity' pair, see the beginning of NE 236.

3. For discussion of the Cartesian view of personal identity, see Wilson (1999: 375–77) and Bobro (2004: 8–9).

4. See also AG 243 and DM 34.

in accounting for 'I' as a moral and religious agent. For Leibniz, 'I' am therefore structured by two distinct kinds of identity, so to speak.⁵

Leibniz's hybrid view of 'I' has been the object of significant controversy. For instance, Margaret Wilson famously maintains that the view contains a serious inconsistency, which indicates that Leibniz fails to harmonize the Cartesian aspects of his view with its Lockean aspects. In "Leibniz: Self-Consciousness and Immortality in the Paris Notes and After", Wilson argues that the following two propositions endorsed by Leibniz entail an inconsistency:

- (1) 'I *am* a particular immaterial substance'.
- (2) 'It is metaphysically possible that I continue as an identical self-consciousness and identical self, independently of this particular substance'. (Wilson 1999: 380)⁶

Concerning (1), Wilson comments, 'there is plenty of evidence in Leibniz's mature writings' that Leibniz 'identified the denotation of 'I' with a particular substance' (Wilson 1999: 377). Moreover, according to Wilson, Leibniz maintains that our self-consciousness of 'I' provides us with an original understanding of the nature of substances identical to us. From these points, Wilson concludes that Leibniz holds that for any 'I', 'self-consciousness must be consciousness of a particular simple substance (the one that is me), and further that it must involve consciousness of the identity, simplicity and substantiality of this entity' (ibid). That is, for Wilson, (1) *also implies that I am the consciousness of this particular substance*. As for textual evidence supporting (1), Wilson regards the following passages as representative:

These souls [rational souls or spirits] are capable of performing reflexive acts, and of considering what is called 'I' [*Moy*], 'substance,' 'soul,' 'mind' in a word, things and truths which are immaterial. (L: 638)

We experience ourselves a multitude in the simple substance, when we find that the least thought which we perceive envelops a variety in its object. Hence everyone who recognizes that the soul is a simple substance should recognize this multitude in the monad. (Mon: 16)

In natural perception and in sensation, it is enough for what is divisible and material and dispersed into many entities to be expressed

5. Przemysław Gut (2017: 100–4) presents a similar account.

6. Wilson's charge of inconsistency was originally presented in her 1976 article; here we cite the version in her 1999 essay collection. Samuel Scheffler (1976: 235–39) presents similar criticism.

or represented in a single indivisible entity or in substance which is endowed with genuine unity. One cannot doubt the possibility of a noble representation of many things in a single one, since our soul provides us with an example of it. (G II.112/LA: 144)

... [B]y means of the soul or form there is a true unity which corresponds to what is called 'I' in us; which could not occur in artificial machines, nor in the simple mass of matter, however organized it may be; which can only be regarded as like an army or herd. ... If there were no true substantial unities, however, there would be nothing substantial or real in the collection. (L: 456)

It is very true that our perceptions or ideas come either from the exterior senses, or from the internal sense, which can be called reflection: but this reflection is not limited to just the operations of the mind, as is said [by Locke.] ... it goes as far as the mind itself, and it is in perceiving [the mind] that we perceive substance. (A VI.vi.14)

The problem is that Leibniz also seemingly endorses (2) in the *New Essays*, i.e., the possibility of a divorce between the two kinds of identity. As already seen, Leibniz comments in the *New Essays* that personal identity presupposes the identity of substance due to the order of nature (NE: 236). But the proviso 'according to the order of things' attached to his comment strongly indicates that it is merely *hypothetically necessary* that our personal identity is conjoined with substantial identity. In other words, the comment seems to admit a *metaphysically possible* situation in which our personal identity is separated from the substance (Wilson 1999: 380).⁷ What makes matters worse is that in the *New Essays* Leibniz seems to say that I can continue as the same self-consciousness independently of this particular substance. For instance, Wilson appeals to the following passages in which Leibniz seems to admit the possibility of the same continuous personal identity being transferred from one soul to another spiritual substance:

Even if God were to change the real identity in some extraordinary manner, the personal identity would remain, provided that the man preserved the appearances of identity—the inner ones (i.e., the ones belonging to consciousness) as well as outer ones such as those consisting in what appears to other people. (NE: 237; our emphasis)

7. More detailed information on the differences between metaphysical necessity and hypothetical necessity is provided in section 3. What is important for our purposes here is only that hypothetical necessity has traditionally been treated as weaker than metaphysical necessity.

I admit that *if God brought it about that consciousnesses were transferred to other souls, the latter would have to be treated according to moral notions as though they were the same*. But this would disrupt the order of things for no reason, and would divorce what can come before our awareness from the truth—the truth which is preserved by insensible perceptions. (NE: 242; our emphasis)

But if it is possible that my personal identity remains the same even when I cease to be identical with this particular substance, it will be difficult to maintain that I am just this immaterial substance. Therefore, Leibniz, in his mature period, seemingly holds an incoherent picture of ‘I’.

Since Wilson’s criticism was raised, various solutions have been proposed to rescue Leibniz from the inconsistency. As we argue below, almost all of the logical options that can be adopted in response to Wilson’s challenge have been developed by Leibniz commentators. However, in our view, there is still work to be done on this issue, since the previously proposed solutions are all deeply problematic. In what follows, we criticize the existing solutions and develop two novel ones. We proceed in the following manner: In section 2, we introduce existing solutions and explain why they are problematic. In section 3, we present a new solution based on Leibniz’s theory of modality. In section 4, we present an alternative solution based on Leibniz’s ambiguous use of the first-person pronoun.

2. Examination of Existing Solutions

This section considers three solutions to Wilson’s conundrum. First, we examine a solution proposed by Nicholas Jolley, namely, that (1) and (2) are, in fact, consistent. Second, we discuss Ezio Vailati’s solution, which maintains that Leibnizians should reject (1). Lastly, we focus on Marc Bobro’s argument that Leibnizians can deny (2).

Before addressing the existing solutions, it is worth noting that Wilson’s charge of inconsistency relies on certain assumptions she does not make explicit. Clarifying these presuppositions can help us better understand both the existing solutions and our novel proposals. Let us therefore briefly outline them in advance: (3) Identity statements (at least those involving names, indexicals, or demonstratives) are necessary; (4) The term ‘I’ in (1) and (2) is unambiguous; (5) The term ‘I’ in (1) and (2) refers to the same thing. Jolley’s solution and our two novel proposals are based on rejecting one of these presuppositions.

2.1. Jolley's solution: Deny the necessity of identity

Nicholas Jolley has suggested that (1) and (2) can be maintained without inconsistency as long as (3) is false, that is, as long as identity statements can fail to be necessary. Jolley reasons as follows: (1) does not have to mean that I am *this* particular immaterial substance (say S_i) with which I am currently identical, for (1) merely states that I am *a* particular immaterial substance. Thus, the first proposition, in itself, allows the possibility that I am merely contingently identical with S_i .⁸ Accordingly, it might be argued that although Leibniz holds that I do obtain certain knowledge about the nature of immaterial substance through my self-consciousness, what I can be aware of is, at best, that I am *an* immaterial substance with general qualities shared by all other souls (e.g., that I am simple, immaterial, etc.), not that I am *this* S_i .⁹ But (1) interpreted in this way is surely compatible with (2), for what (1) demands according to this interpretation is just that 'I' should somehow take root in a soul, not that I am always identical with S_i . In short, Jolley's solution is to claim that, for Leibniz, (3) is, in fact, false, in that Leibniz allows contingent identity.

The Leibniz–Arnauld correspondence (1686–1690) appears to provide textual support for Jolley's solution, for Leibniz comments in his July 1686 letter as follows: 'It is not enough for understanding the nature of myself, that I feel myself to be a thinking substance, one would have to form a distinct idea of what distinguishes me from all other possible minds; but of that I have only a confused experience' (G II.52/LA: 59). Here, Leibniz explicitly admits that we cannot form a distinct idea of who we are. In other words, it seems that the most I can know about myself is that I am an immaterial substance, not that I am S_i , because I can, at best, have only a confused knowledge about what distinguishes

8. To our knowledge, no published article explicitly defends this option. Bobro reports, however, that Jolley said in private correspondence that he was attracted to this way of understanding the problematic comments in the *New Essays* (Bobro 2004: 58, fn. 28). In his *The Light of the Soul*, Jolley indeed appeals to the distinction between 'the idea of the soul in general and the idea of an individual soul' when interpreting Leibniz's position on self-knowledge based on a comparison among Leibniz's, Malebranche's, and Locke's views on the issue (Jolley 1990: 176). There, Jolley emphasizes that this distinction can play a key role in resolving the tension between Leibniz's comment that we do not have clear knowledge of our soul and his further claim that we do have some certain knowledge of the nature of our soul. Thus, although the issue on which Jolley himself focused is slightly different from the main topic of this paper, there is a good reason to credit Bobro's report that Jolley was inclined to adopt a similar strategy to provide an answer to Wilson's challenge. For these reasons, we refer to the proposal described in this sub-section as 'Jolley's solution'. Wilson also considers this strategy as a possible response to her challenge (Wilson 1999: 381).

9. Wilson maintains that the assumption that I do not have a definitive consciousness of myself as this specific substance is necessary to make (1) and (2) consistent (Wilson 1999: 381). Bobro also treats this assumption as one of the necessary conditions for any plausible interpretation on which (1) and (2) are, in fact, consistent (Bobro 2004: 46).

me from other substances. If so, self-consciousness is merely a consciousness of *an* immaterial substance, not of S_1 . Therefore, based on LA 59, one might argue that (3) is false and that (1) and (2) are, in fact, consistent.

Jolley's proposal is flawed for several reasons. First, it is unclear whether LA 59 really supports Jolley's suggestion. There, Leibniz only says that to understand the nature of myself I must form a *distinct* idea of who I am, which is beyond the scope of my understanding. However, to make this claim support Jolley's interpretation, we need an additional assumption, namely, that a distinct knowledge of my nature is required to be aware of whether I am identical with S_1 . But this premise might seem too demanding. For instance, we can readily distinguish yellow from other colors even if we do not have complete knowledge of the nature of that color. So why can't we assume that I can distinguish S_1 from S_2 without a distinct knowledge of my nature?

In his letter to Thomas Burnett (1699), Leibniz indeed argues that we do not need a distinct knowledge of the nature of a particular substance to recognize it or distinguish it from others:

Now, if I dared to mix my thoughts in with the thoughts of these excellent men [i.e., Locke and Edward Stillingfleet], I would distinguish between clear and distinct, as I did elsewhere in the *Acts* of Leipzig. I call an idea clear when it is sufficient for recognizing a thing, as when I remember a color well enough to recognize it when it is brought to me; but I call an idea *distinct* when I conceive its conditions or requisites, in a word, when I have its definition, if it has one. Thus I do not have a distinct idea of all colors, being often required to say that it is a something-I-know-not-what that I sense very clearly, but cannot explain well. And similarly, I believe that we have a clear idea, but not a distinct idea, of substance, which arises, in my opinion, from the fact that we who are substances have an internal sensation [*sentiment*] of it in ourselves. (AG: 287)

As can be seen here, Leibniz defines a distinct idea as an idea by which we can conceive the 'requisites' of the idea (or the object of the idea).¹⁰ But as Leibniz himself points out, we can sometimes readily recognize an entity and distinguish it well from others even if we do not possess a distinct idea of it—as in the case of color. In Leibniz's view, we have, in such cases, a *clear but indistinct* idea. Leibniz then explicitly maintains that our idea of a substance obtained through self-consciousness is an example of a clear but indistinct idea. This comment is compatible with the reading that I can readily recognize the substance with which I am

10. For Leibniz, a requisite, or prerequisite, of X is such that if it is not posited, then X is not given (A VI.2.483). A distinct idea of an individual substance containing its requisites would be given by its complete individual concept.

identical and distinguish it from other substances. In other words, AG 287 can be interpreted as allowing that I can have a 'clear' recognition that I am this specific substance, S_i , although I may not have a 'distinct' understanding of why I, as S_i , am different from other substances. The Leibniz–Arnauld correspondence (LA: 59) thus does not provide decisive support for Jolley's suggestion.¹¹

Moreover, Jolley's solution crucially relies on the assumption that Leibniz rejects (3), namely, that he allows for contingent identity, for if migration of consciousness is possible, the same 'I' might be identical with different substances. However, Leibniz emphasizes in multiple texts that identity statements are necessary. For instance, in "Necessary and Contingent Truths" (1686), Leibniz comments that 'an *absolutely necessary* proposition is one which can be resolved into identical propositions, or, whose opposite implies a contradiction' (MP: 96). Likewise, in the *New Essays*, Leibniz regards truths about identity as necessary: 'Truths of reason are necessary, and those of fact are contingent. The primary truths of reason are the ones to which I give the general name "identities"' (NE: 361). Such comments suggest that if I am identical with S_i , then it is metaphysically necessary that I am identical with S_i : For, generally, $(A=B) \rightarrow \Box(A=B)$. Therefore, Jolley's proposal is not promising.

2.2. Vailati's solution: Reject (1)

Ezio Vailati (1985) argues that we should reject (1), that 'I *am* a particular immaterial substance'. In other words, Vailati holds that for Leibniz, 'I' as a person can be detached from *this* particular substance that currently underlies my apparent identity. Vailati's rejection of (1) suggests that he accepts that (1) and (2) are unambiguous [i.e., presupposition (4)] and that the first-person pronouns in (1) and (2) refer to the same thing [i.e., presupposition (5)], while acknowledging the necessity of identity claims [i.e., presupposition (3)], unlike Jolley. For one should reject either (1) or (2) only when presuppositions (3), (4), and (5) are

11. We intentionally use cautious language in presenting our interpretation of AG 287 because, as an anonymous referee points out, the passage may allow for an alternative interpretation. For instance, some might propose that the passage merely suggests that our sensations provide a sufficiently clear idea of substance (in general), allowing us to distinguish substances from one another (e.g., through the idea that no two substances share all the same properties). However, this does not imply that I have a clear idea of *my* substance as distinct from others. Given the possibility of this alternative interpretation, we refrain from claiming that AG 287 *entails* that I have a clear idea of *my* substance *on any interpretation*. Instead, we claim only that the textual argument for Jolley's solution based on LA 59 is not decisive. Bobro (2004: 47) also offers a similar but more succinct criticism of Jolley's interpretation. As far as we are aware, the criticism we make of Jolley's consistency proposal in the next paragraph is original and robust enough to show that Jolley's solution is problematic regardless of how LA 59 is interpreted.

simultaneously endorsed. To justify his position, Vailati provides two reasons—one textual and the other philosophical.

Begin with the textual reason. Vailati emphasizes that ‘Wilson does not provide any textual evidence for the claim that in the *New Essays* Leibniz holds [(1)], although there is little doubt that Leibniz holds [(1)] in the *Discourse on Metaphysics*, written almost twenty years before’ (Vailati 1985: 36–37). In short, Vailati maintains that as a result of Leibniz changing his mind at some point in his philosophical career, ‘according to Leibniz’s theory of personal identity in the *New Essays* [(1)] is false’ (Vailati 1985: 37).

This textual reasoning is weak. It is true that Wilson does not provide any passages from the *New Essays* to support (1). However, Wilson’s first piece of textual evidence comes from “Principles of Nature and of Grace” (1714), written after the *New Essays* (1704), and her last piece of evidence comes from a 1698 manuscript containing Leibniz’s reflections on the Second Book of Locke’s *Essay* (Wilson 1999: 386, fn. 30). Therefore, there is no good reason to think that Leibniz changed his view when he wrote the *New Essays*. Moreover, as Bobro (2004) points out, some passages from the *New Essays* can be interpreted as supporting (1). For instance, Leibniz says, ‘one can rightly say that they [i.e., substantial beings] remain perfectly “the same individual” in virtue of this soul or spirit which makes the *I* in substances which think’ (NE: 232). Here, Leibniz holds that this ‘soul or spirit’ is what makes the ‘*I*’. Similarly, Leibniz comments, ‘I would rather say that the *I* and the *he* are without parts, since we say, quite correctly, that he continues to exist as really the same substance, the same physical *I*’ (NE: 238). In this passage, Leibniz clearly states that ‘*I*’ is a substance without parts, i.e., a soul. Therefore, Vailati needs to present an additional argument to show that these passages should not be read as supporting (1), considering his strong claim about the changes in Leibniz’s view.¹²

What of the philosophical reason? Vailati’s argument relies crucially on the so-called ‘thinking machine’ passage. To set the context: In the *Essay*, Locke takes an agnostic attitude toward whether a thinking machine—a system of matter to which the faculty of thinking is superadded by divine miracle—is impossible:

We have the *Ideas of Matter and Thinking*, but possibly shall never be able to know, whether any mere material Being thinks, or no; it being impossible for us, by the contemplation of our own *Ideas*, without revelation, to discover, whether Omnipotency has not given to some Systems of Matter fitly disposed, a power to perceive and think, or else joined and fixed to Matter so disposed, a thinking immaterial Substance: It being, in re-

12. For Bobro’s criticism of Vailati’s solution based on textual evidence from the *New Essays*, see Bobro (2004: 45).

spect of our Notions, not much more remote from our Comprehension to conceive, that GOD can, if he pleases, superadd to Matter a Faculty of Thinking, than that he should superadd to it another Substance, with a Faculty of Thinking. (E IV.3.6)

Regarding the possibility of a thinking machine, Leibniz holds that although such a machine is metaphysically possible in virtue of divine omnipotence, it is precluded by the order of nature:

You [i.e., Locke] seem to hold, sir, that this apparent identity could be preserved in the absence of any real identity. Perhaps that could happen through God's absolute power; ... *If a man could be a mere machine and still possess consciousness, I would have to agree with you, sir; but I hold that that state of affairs is not possible—at least not naturally.* (NE: 236; our emphasis)

In the cited passage, Leibniz treats a thinking machine as an example of a case in which an agent's apparent identity is separated from its real identity. The reason is that while the thinking machine might have self-consciousness, it cannot be treated as a substance because it consists of divisible matter, which lacks unity. Leibniz's comment that this machine is 'not possible—at least not naturally' suggests that he leaves room for the metaphysical possibility of a thinking machine, even if such a machine is hypothetically impossible.

The gist of Vailati's philosophical argument is that there are good reasons to attribute personal identity to a thinking machine because this machine would be endowed with self-consciousness if it were to be created by God. The machine would self-reflect, have apperceptions and memories, and treat itself as 'I' just as ordinary souls do. However, while the machine would enjoy the same personality as ordinary 'I's, it is not an immaterial substance, nor can it possess direct knowledge of the very substance of its mind, simply because there is no substance at bottom. A thinking machine can, at best, possess a *representation* of its substantial self: 'M [i.e., the thinking machine] perceives the representation of its substantial self while there is no substantial self to be perceived' (Vailati 1985: 39). To sum up, Vailati maintains that the possibility of a thinking machine gives us a good reason to reject (1), for not all entities suited for the attribution of 'I' are souls.

However, in our view, this argument is not successful, for it crucially relies on the assumption that it is justifiable to ascribe personality to the thinking machine. Vailati indeed assumes that the thinking machine would become a moral agent if God were to create it: '[A] thinking machine, if God decided to produce one, would be a legal and moral subject for it could think, make plans, act, remember and feel such things as guilt, pain, and pleasure' (Vailati 1985: 39). However, as Bobro (2004) points out, Leibniz provides good reason to be skeptical of this

assumption. In a letter to Damaris Masham (1704), Leibniz comments that since matter is inapt to produce the modifications of a soul, God would have to miraculously sustain its power of thinking if a thinking machine exists:

[T]he illustrious Locke maintained in his excellent *Essay* and his writings against the late Bishop of Worcester that God can give matter the power of thinking, because he can make everything we can conceive happen. But then matter would think only by a perpetual miracle, since there is nothing in matter in itself, that is, in extension and impenetrability, from which thought could be deduced, or upon which it could be based. ... *God, in the case of thinking matter, must not only give matter the capacity to think, but he must also maintain it continually by the same miracle, since this capacity has no root [racine], unless God gives matter a new nature.* But if one says that God gives matter this new nature or the radical power to think, since that power is maintained by itself, he would simply have given it a thinking soul, or else something that differs from a thinking soul only by name. (AG: 290; our emphasis)

Leibniz's comment that the capacity of thinking has no root in a thinking machine suggests that the machine is not a spontaneous agent of its thoughts, for 'there is nothing in matter in it self ... from which thought could be deduced, or upon which it could be based'.¹³ Accordingly, in AG 290 Leibniz explains the need for perpetual miracles based on matter's complete lack of spontaneity in giving rise to its thoughts. However, for Leibniz, spontaneity is one of the necessary conditions for freedom (T: 288). It follows that a thinking machine is, by nature, not a free agent, due to its lack of spontaneity in its thoughts. If so, there is no reason to hold that the thinking machine would enjoy the status of a person.¹⁴ Therefore, neither Vailati's textual reasoning nor his philosophical reasoning solves the problem of inconsistency.

2.3. Bobro's solution: Reject (2)

Marc Bobro (2004) maintains that the passages in which Leibniz seemingly supports (2) (i.e., that the separation between the two kinds of identity is possible)

13. According to Leibniz's doctrine of spontaneity, all finite substances are causally insulated from any external influence from other finite substances (e.g., Mon: 11; T: 290; AG: 279). This insulation from external influence implies that a monad can completely produce or determine its succeeding states based on the previous state (LA: 51).

14. For Bobro's interpretation of the 'thinking machine' passage, see Bobro (2004: 60–68). Bobro also maintains that Leibniz's system does not allow the possibility that the thinking machine enjoys genuine moral agency.

can be interpreted as not in fact supporting (2). Bobro's rejection of (2), of course, indicates that he endorses (3) the necessity of identity claims, (4) the unambiguity of 'I' in (1) and (2), and (5) the claim that 'I' in (1) and 'I' in (2) refer to the same thing. To support his position, Bobro first considers the following three passages as a group, some of which we have already seen in our previous discussion:

- (a) You [i.e., Locke] seem to hold, sir, that this apparent identity could be preserved in the absence of any real identity. Perhaps that could happen through God's absolute power ... If a man could be a mere machine and still possess consciousness, I would have to agree with you, sir; but I hold that that state of affairs is not possible—at least not naturally. (NE: 236)
- (b) I admit that if God brought it about that consciousnesses were transferred to other souls, the latter would have to be treated according to moral notions as though they were the same. But this would disrupt the order of things for no reason ... (NE: 242)
- (c) I acknowledge that if all the appearances of one mind were transferred to another, or if God brought about an exchange between two minds by giving to one the visible body of the other and its appearances and states of consciousness, then personal identity would not be tied to the identity of substance but rather would go with the constant appearances, which are what human morality must give heed to. (NE: 244)

According to Bobro, 'Leibniz makes a two-fold concession to Locke in these passages, *neither* of which commit Leibniz to [(2)]' (Bobro 2004: 52): First, Leibniz acknowledges the metaphysical possibility of a thinking machine and migrating consciousnesses. Second, 'he admits that we ought to treat that which possesses apparent identity ... as persons or morally responsible agents' (ibid.). Why don't these concessions, then, mean that Leibniz endorses (2)?

As Bobro argues, if we strongly adhere to the first proposition of Wilson's puzzle—that 'I' refers to an immaterial substance—we should not acknowledge any personal identity separate from substantial identity. That is, even if there is a case in which an entity seems to enjoy the status of a person without a substantial basis, proponents of (1) must maintain that, in such a case, the entity can at best have 'apparent' identity, not genuine personal identity. However, human morality is not sensitive to the subtle metaphysical differences between the appearance of personal identity and a person. Moreover, as Leibniz points out in passage (c), it is 'the constant appearances', not the real identity, that 'human morality must give heed to' (NE: 244). In sum, Bobro maintains that although (a)–(c) might support the claim that we ought to treat individuals with apparent identity without substantial identity as persons, they do not imply that apparent identity constitutes genuine personal identity (Bobro 2004: 52).

Bobro also considers a fourth passage that seemingly supports (2). Before discussing it, let us briefly comment on Bobro's interpretation of (a)–(c). We do not take issue with his treatment of (a) and (b). However, it is unclear whether Bobro's solution can apply smoothly to (c). There, Leibniz holds that if migration of consciousness were to happen, 'personal identity would not be tied to the identity of substance but rather would go with the constant appearances' (NE: 244). This comment seemingly implies that if migration were to occur, personal identity would not be separated from apparent identity and instead would 'go with' the transferred appearance. That is, (c) seems to entail that in the case of migration, the transferred consciousness C_i would constitute the same person as the original consciousness C_o . However, this consequence is identical to Wilson's second proposition.

Bobro encounters a similar problem when dealing with the fourth passage (d), a problem that suggests that it is a mistake to group (c) with (a) and (b) instead of with (d). Bobro admits that his strategy does not work for (d):

(d) Even if God were to change the real identity in some extraordinary manner, the personal identity would remain, provided that the man preserved the appearances of identity. (NE: 237)

Here, Leibniz strikingly uses the verb *remain*, which indicates that even if C_o were to experience a switch of souls, it would retain personal identity insofar as it preserved its appearance. Therefore, (d), as well as (c), is in tension with Bobro's strategy, which separates apparent identity from personal identity.

Since his strategy is no longer available for interpreting passage (d), Bobro tries to rely on Benson Mates's (1986) interpretation of that passage: 'If Mates is right, then this is how we *should* read the specious persons passage [i.e., NE 237]' (Bobro 2004: 53). Concerning (d), Mates suggests the following reading:

At A.6.6.237 [i.e., NE 237], it looks as if Leibniz is countenancing the possibility that X at t might be morally identical with Y at t' even if they were not really identical, but I think that in that passage the clause '... if God were to change the real identity in some extraordinary manner' need only mean '...if the temporal development of the soul in question contained a discontinuity'. (Mates 1986: 145, fn. 24)

Mates maintains that NE 237 can be interpreted as not being committed to the possibility of the migration of consciousness. In Mates's view, (d) only concedes the possibility of a discontinuity in the series of perceptual states temporally developed in the soul—i.e., the possibility that the perception following the

preceding perception p_o is not p_1 (i.e., the perception produced by p_o in a continuous manner) but p_1^* , which is entirely disconnected from the previous state. Will Mates's reading save Leibniz from the charge of inconsistency?

Unfortunately, Bobro's proposal based on Mates's reading does not solve the problem. After introducing Mates's interpretation, Bobro focuses on showing that perceptual discontinuity can lead to a *change* in the real identity: 'However, there is a way in which God could change a substance's real identity without destroying that substance once and for all; for example, if its states violated the principle of continuity' (Bobro 2004: 54). The same point is repeated a little later: 'Would this disruption of perceptual continuity change the real identity? ... surely a substance would be genuinely altered if its states were no longer continuous' (Bobro 2004: 54–55). However, it is unclear how this point supports rejecting (2). Following Bobro, suppose that (d) contains Leibniz's concession of the possibility of perceptual discontinuity. On this supposition, the original passage (d) should be read as (e):

- (e) Even if the temporal development of the soul in question contained a discontinuity, the personal identity would remain, provided that the man preserved the appearances of identity.

However, if the discontinuity of perceptual development leads to a change in real identity—as Bobro suggests—(e) says, at best, that even if there were to be a change in real identity, the personal identity would remain. But this is only a different way of expressing (2). So Bobro's reading still allows for the separation between the two kinds of identity, although the divorce, in this case, does not occur through a thinking machine or the migration of the soul. Thus, Bobro's reading of (d) is problematic: The very reason he gives for rejecting (2) becomes a reason to accept it.

Bobro might try to avoid this criticism by accepting that perceptual discontinuity does not necessarily lead to a change in substantial identity. But, even if he takes this tack, there still remains a problem: Why should we assume that the if-clause in NE 237 concerns perceptual discontinuity in particular? As we have seen in the other passages cited by Bobro, Leibniz consistently keeps in mind the possibility of a thinking machine and the migration of consciousness in his account of personal identity. The consideration of the broader context suggests that it is much more natural to interpret God's 'some extraordinary manner' mentioned in (d) as related to a thinking machine or the migration of consciousness. Thus, unless a clear reason is given as to why NE 237 should be interpreted in connection with the issue of perceptual continuity, Wilson's backers will treat Bobro's suggestion as an ad hoc solution.

3. The Modal Solution

Thus far, we have shown that the existing solutions are all unsuccessful. We will now present a novel ‘modal’ solution. The primary feature of this solution lies in the denial of (5)—that ‘I’ in (1) and ‘I’ in (2) refer to the same thing. According to the modal solution, (1) and (2) are compatible because the referent of ‘I’ in (1) differs from the referent of ‘I’ in (2) due to Leibniz’s conception of the distinction between two modalities (i.e., metaphysical and hypothetical). By rejecting (5), this solution preserves the consistency between (1) and (2) without rejecting (3) or the necessity of identity statements. At the same time, this solution maintains that there is no ambiguity in the term ‘I’ in (1) and (2), provided that we have a clear grasp of Leibniz’s view of modality. While Bobro’s solution fails to provide a consistent way of dealing with the four problematic passages (a)–(d), our modal solution represents a more systematic approach to the issue.

As we have seen, Leibniz continuously emphasizes in the *New Essays* that although the divorce between the two kinds of identity is possible, the laws of nature determine that personal identity presupposes substantial identity (NE: 236, 242, 245). As Larry Jorgensen (2019) rightly points out, for Leibniz this physical determination relationship between the two kinds of identity is an instance of *hypothetical necessity*.¹⁵ In other words, although the divorce between the two kinds of identity is metaphysically possible, that separation is hypothetically impossible.

How, then, can Leibniz’s distinction between the two necessities help us solve Wilson’s conundrum? For Leibniz, there is no difference between metaphysical and hypothetical necessities in terms of their modal strength.¹⁶ Instead, the difference stems from their structural differences in obtaining the same degree of modal strength: While something metaphysically necessary obtains its modal strength from its essence, something hypothetically necessary relies on a source external to its essence to achieve its necessity.¹⁷ For instance, in “On Free-

15. See Jorgensen (2019: 248); see also NE 178–79 and DM 13. Scheffler (1974: 237) alternatively relies on the notion of moral necessity to capture the relationship between the two kinds of identity (e.g., it is morally necessary that personal identity presupposes substantial identity), but the difference between moral and hypothetical necessity is not an important issue in this context. The main reason for that is that in Leibniz’s view the order of nature for a certain world also involves the moral element, i.e., divine consideration of the best (NE: 179).

16. Leibniz makes this point in *Confessio Philosophi* (Fall 1672–Winter 1672/73?). See CP 55–57, where Leibniz explains the differences between the two necessities by making an analogy with the structural differences between true singular propositions and universal propositions. There, Leibniz makes it clear that both propositions are equal in their truth value, which is likened to necessity or modal strength.

17. Adams presents a similar interpretation: ‘In Leibniz’s conception of hypothetical necessity, the absolute necessity or contingency of the antecedent is no more important than the externality of the antecedent to the consequent. *What follows necessarily from what is necessary through itself is*

dom, Fate, and the Grace of God" (1686/87?), Leibniz distinguishes two kinds of necessity as follows:

But now, something following infallibly from what has been posited, and hence being hypothetically necessary, is entirely different from demonstrating something—without any supposition of existing things—solely from the necessity of essences or from terms or ideas, such that its contrary implies contradiction. (A VI.4.1598/LGR: 256–57)

That is, while something metaphysically necessary is necessary due to its essence, something hypothetically necessary is necessary because of certain posited things external to its essence. Similarly, in appendices to the *Theodicy*, Leibniz distinguishes metaphysical necessity—described as that which 'exists in that which is essential' (Hug: 384) or is 'necessary of itself' (Hug: 397)—from hypothetical necessity, which arises because of 'anterior reasons' or 'as a result of the supposition that this or that has been foreseen or resolved, or done beforehand' (ibid.). In short, *the reason hypothetical necessity differs from metaphysical necessity is not that it allows a possible world in which the relevant state of affairs does not obtain, but that it needs to rely on an external source to obtain its necessity, or for its being true in all possible worlds.*

If so, we cannot straightforwardly apply contemporary modal analysis to hypothetically impossible states of affairs: It may be that, for Leibniz, when p is hypothetically impossible, ' p is metaphysically possible' does not imply that there is a possible world in which p occurs. This reading can also be supported by textual evidence. Leibniz suggests sentences such as 'Caesar becomes a dictator' as an example of a metaphysically contingent but hypothetically necessary statement (DM: 13). If so, the following statement is true:

(f) It is metaphysically possible that Caesar does not become a dictator.

However, there is probative textual evidence that, for Leibniz, (f) does not mean that there is a possible world in which an individual identical to Caesar does not become a dictator. For example, at the end of the *Theodicy* (1710), in discussing the case of Sextus who is wicked and unhappy in the actual world (T: 413), Leibniz states that alternative possibilities in which Sextus is 'very happy and noble' or 'content with a mediocre state' do not involve 'absolutely the same Sextus ... (that is not possible, he carries with him always that which he shall be) but several Sextuses resembling him' (T: 414). If so, what would make (f) true is

certainly necessary by necessity of the consequent, in the traditional sense. It is not necessary through itself, however, but only hypothetically necessary, and contingent, in Leibniz's sense, if the antecedent from which it follows is external to it, and not contained in its own nature' (Adams 1994: 17–18; our emphasis).

not the existence of a possible world in which *Caesar* does not become a dictator, but rather the existence of a possible world in which *Caesar**—something like a counterpart of *Caesar*—does not become a dictator.¹⁸

Now, let us return to the original question. Wilson's second proposition is as follows:

- (2) 'It is metaphysically possible that I continue as an identical self-consciousness and identical self, independently of this particular substance'. (Wilson 1999: 380)

Let us suppose that (2) can be simplified as follows:

- (2)' It is metaphysically possible that person P_o can continue independently of this particular substance S_o .

If (2)' implies that there is a possible world in which P_o can continue even if it is separated from S_o in which it is rooted, (2)' indeed contradicts (1), the claim that this person is just the same as S_o . However, Leibniz highlights that the separation between the two kinds of identity is hypothetically impossible, and we now know that, for Leibniz, hypothetically necessary statements should sometimes be analyzed differently than their appearance suggests. For instance, we cannot rule out the possibility that the truth condition of (2)' is as follows:

18. To be fair to Wilson, it should be noted that she is highly skeptical of applying contemporary possible world semantics to interpret Leibniz's modal theory. In particular, in "Possible Gods" (1979) Wilson criticizes Hidé Ishiguro's (1972) and Fabrizio Mondadori's (1973) attempts to interpret Leibniz's discussions of counterfactuals using David Lewis's (1968) counterpart theory. Her criticism proceeds in two main directions. First, she argues that applying counterpart theory uniformly to all modal statements that Leibniz is interested in leads to the absurd conclusion that there are counterparts of God in Leibniz's system (Wilson 1999: 411–12). Second, she questions whether Leibniz has the tools to ensure that there is only one closest possible world, a crucial requirement for determining the truth value of counterfactuals. For example, if there are multiple closest worlds in which Sextus* does not commit a sin towards Lucretia*, and whether Sextus* lives a happy life varies among these closest worlds, the truth value of the counterfactual cannot be determined (Wilson 1999: 410–11). However, we do not find Wilson's criticisms decisive, as Michael Griffin offers a compelling response to them. Griffin argues that there are good reasons to attribute to Leibniz a counterpart semantic theory for counterfactuals concerning *created individuals* (Griffin 1999: 338). One reason is that there is textual evidence suggesting that *Leibniz does have the tools to disarm Wilson's second criticism* (Griffin 1999: 329–33). Of course, it is possible that Griffin is mistaken, in which case it might be better to adopt the ambiguity solution (see below) over the modal solution, as the latter commits Leibniz to a counterpart semantics of some sort. Nonetheless, given the strong case that Griffin makes, one cannot simply dismiss the claim that Leibniz's counterfactuals should be analyzed in counterpart-theoretical terms as implausible or anachronistic. One would need stronger arguments responding to Griffin.

(2)* 'It is metaphysically possible that P_o can continue independently of this particular substance S_o' is true iff there is a possible world in which P_o^* can continue independently of S_o^* .

If the truth condition of (2)' is (2)*, there is no contradiction between (2)' and (1). From (2)' and (2)* we can, at best, conclude that there is a possible world in which the counterpart of person P_o can continue independently of the counterpart of substance S_o . Of course, as far as we know, Leibniz himself does not provide any official account concerning how counterpart relationships are determined. Therefore, the details concerning how counterparts S_o^* and P_o^* should be understood can become a matter for further interpretation. For instance, while some might hold that a thinking machine can also be a counterpart of S_o due to its appearance being indistinguishable from S_o , others might be reluctant to treat the machine as one of S_o 's counterparts due to its lack of substantiality.

However, what is crucial in the context of our argument is only that *the counterpart of P_o referred to in (2)*, whatever it may be, cannot enjoy the status of a person as the original P_o does*. To see why, consider a possible world in which P_o^* is detached from S_o^* and transferred to S_1^* at t_n through God's miraculous intervention. In this counterfactual scenario, P_o^* , after its attachment to S_1^* at t_n , cannot be regarded as a person for the following reasons. First, suppose S_1^* —the counterpart of substance S_1 —is not a substance (e.g., it is a thinking machine resembling S_1). In this case, it is evident that P_o^* cannot be treated as a person after its attachment to S_1^* because S_1^* , as a thinking machine, lacks the ability to produce perceptions and memories in P_o^* . In other words, P_o^* , in this case, lacks a ground for its mental contents. Next, suppose S_1^* is a substance. One might argue that perhaps S_1^* could provide a new basis for the contents of P_o^* . However, this cannot be the case considering Leibniz's view on spontaneity. For Leibniz, spontaneity means that the subsequent state of a substance is fully determined by its preceding state: '[E]very present state of a substance occurs to it spontaneously and is only a consequence of its preceding state' (LA: 51). Because of this tight connection between preceding and succeeding states, God can infer the entire history of a substance even from its single state (e.g., AG: 41; Mon: 22). Now, if S_1^* were to provide a ground for perceptions and memories contained in P_o^* after t_n (after P_o^* 's attachment to it), the spontaneity doctrine would require that the preceding state contained in S_1^* at t_{n-1} (prior to P_o^* 's attachment) be identical to the prior state of S_o^* at t_{n-1} .¹⁹ This would imply that S_1^* and S_o^* share the same

19. To avoid this consequence, one might argue that there is a possibility for the same subsequent state to arise from different preceding states. However, as previously pointed out, Leibniz asserts that God can read the entire history of a substance, even from a single fragment of its content (AG: 41). This is possible because the present state itself already contains a trace left by the unique past that this substance has experienced: '[T]he present is filled with the future and is laden

states through the past, present, and future. Given Leibniz's endorsement of the identity of indiscernibles, this sharing of entire states and history leads to the conclusion that S_o^* and S_i^* are in fact identical—contradicting the initial presupposition that they are distinct. Thus, even when S_i^* is a substance, P_o^* attached to S_i^* cannot be regarded as a person for lack of a spontaneous source for its content. Since, as assumed in (2)*, we suppose from the outset that P_o^* can exist independently of S_o^* , it follows that P_o^* is not a person from the beginning. For if P_o^* were to be a person, it would no longer survive after its attachment to S_i^* .²⁰ Consequently, although (2)' employs the term 'person,' our analysis of its truth condition shows that it does not, in fact, concern persons. Hence, Leibniz can be acquitted of the charge of inconsistency between (1) and (2).

Let us then finish this section by testing the modal solution against individual passages. First, consider passage (d), for which Bobro failed to provide a plausible reading:

- (d) Even if God were to change the real identity in some extraordinary manner, the personal identity would remain, provided that the man preserved the appearances of identity. (NE: 237)

In our view, leaving out the reference to God, (d) can be innocuously translated as follows:

- (d') It is metaphysically possible that person P_o remains even if P_o 's basis is changed from substance S_o to S_i .

According to our solution, the truth condition of (d)' is as follows:

with the past' (AG: 296). If the same subsequent state is compatible with different series of the past, it is somewhat difficult to understand how a single present state could be described as laden with one substance's unique past. Therefore, considering the textual evidence, there is good reason to hold that Leibniz prefers a strong reading of the spontaneity doctrine—i.e., that one subsequent state/effect can come only from a unique series of past states.

20. For this reason, the option interpreting P_o^* as a person similar to the original P_o is excluded. Even when some entity indistinguishable from P_o^* seems to remain after the detachment, what truly is there is nothing more than a marionette manipulated by brute divine power. By similar reasoning, we can additionally show that interpreting the truth condition of (2)' without appealing to the counterparts of persons is also not a viable option. That is, some might argue that while the notion of counterpart can be applied to substances, it is uncertain whether the same notion can be applied to persons. If this is the case, then the truth condition of (2)' would be that there is a possible world in which person P_o can continue regardless of S_o^* . However, due to the close connection between spontaneity and freedom within Leibniz's system, it again turns out that there is no such world. Of course, our argument crucially relies on the assumption that an entity without spontaneity and freedom cannot be regarded as a person. For those who do not want to accept this premise, we offer a different solution in the next section.

- (g) 'It is metaphysically possible that P_0 remains even if P_0 's basis is changed from substance S_0 to S_1 ' is true iff there is a possible world in which P_0^* remains even if P_0^* 's basis is changed from S_0^* to S_1^* .

But (g) merely implies that there is a world in which an entity similar to P_0 can survive even after its basis is changed from S_0^* to S_1^* . However, there is no reason to hold that this counterpart enjoys the same moral status as the original person P_0 : it is a free-floating being that can survive even when its ground is radically changed, and so its perceptions and memories are not rooted in any specific substance, whether S_0^* or S_1^* . Therefore, (d) is consistent with (1).

Next, let us consider passages (b) and (c):

- (b) I admit that if God brought it about that consciousnesses were transferred to other souls, the latter would have to be treated according to moral notions as though they were the same. But this would disrupt the order of things for no reason ... (NE: 242)
- (c) I acknowledge that if all the appearances of one mind were transferred to another, or if God brought about an exchange between two minds by giving to one the visible body of the other and its appearances and states of consciousness, then personal identity would not be tied to the identity of substance but rather would go with the constant appearances, which are what human morality must give heed to. (NE: 244)

In (b) and (c), Leibniz seems to maintain that migration of personal identity is possible. In other words, it looks like Leibniz endorses the following proposition (h):

- (h) It is metaphysically possible that P_0 can be transferred from S_0 to S_1 .

However, the truth condition of (h) is only that there is a world in which P_0^* can be transferred from S_0^* to S_1^* . Therefore, (b) and (c) do not make an argument substantially different from (d) in that they all maintain that an entity similar to P_0 can be separated from the entities resembling S_0 and S_1 . If there is an additional point that Leibniz adds in (b) and (c), it is only that we should morally treat the counterpart of P_0 as the original person P_0 .

Lastly, let us proceed to (a), the 'thinking machine' passage:

- (a) You [i.e., Locke] seem to hold, sir, that this apparent identity could be preserved in the absence of any real identity. Perhaps that could happen through God's absolute power ... If a man could be a mere machine and still possess consciousness, I would have to agree with you, sir; but I hold that that state of affairs is not possible—at least not naturally. (NE: 236)

In our view, concerning (a), Leibniz may be read as making the following two points:

- (i) It is metaphysically possible that P_o is preserved without substance S_o .
- (j) (This is because) It is metaphysically possible that P_o is (instead) attached to machine M_o .

But according to our general solution, (i) and (j) only imply (k) and (l), respectively:

- (k) There is a possible world in which P_o^* is preserved without S_o^* .
- (l) (This is because) There is a possible world in which P_o^* is attached to M_o^* .

In short, Leibniz only acknowledges in (a) that there is a possible world in which an entity similar to person P_o is preserved without the counterpart of substance S_o , on the grounds that there is a possible world in which P_o^* is attached to M_o^* . To provide a clearer interpretation, we could maintain that Leibniz here admits that an entity similar to person P_o —but which is not a person in itself—can be attached not only to immaterial substances resembling S_o but also to machines similar to our world's M_o . Since (a) interpreted in this way is also about the counterpart of person P_o and not P_o itself, we can conclude that our general solution works well in dealing with the four problematic passages (a)–(d), thereby rescuing Leibniz from the charge of inconsistency.

4. The Ambiguity Solution

The ambiguity solution, unlike all other solutions, denies (4), that (1) and (2) are unambiguous. The ambiguity is such that on one reading (1) is true but (2) is false, whereas on another reading (2) is true but (1) is false. In other words, there is no single interpretation on which both (1) and (2) are true. To see how such ambiguity arises in (1) and (2), let us return to the two propositions that Wilson (1999) ascribes to Leibniz and takes to be mutually inconsistent:

- (1) I am a particular immaterial substance.
- (2) It is metaphysically possible that I continue as an identical self-consciousness and identical self, independently of this particular substance.

In the previous section, we argued that according to Leibniz the truth conditions for (2) invoke counterparts of persons and souls. In this section, we point out that, even if this were not the case, Leibniz still has a way of avoiding contradic-

tion. The reason is that the word 'I' is ambiguous: sometimes 'I' refers to a soul and sometimes 'I' refers to a person (or moral agent).

Recall the following passage from the *New Essays*:

I would rather say that the *I* and the *he* are without parts, since we say, quite correctly, that he continues to exist as really the same substance, *the same physical I*. (NE: 238; our emphasis)

It should be clear that Leibniz here identifies the substance or soul with what he calls 'the physical I'. This is in keeping with Leibniz's use of 'physical identity' to refer to identity of substance. Thus, for example, Leibniz says that the '*incessancy of...soul...preserve[s] real, physical identity*' (NE: 236). It is also in keeping with other language that he ties explicitly both to the term 'real (physical) identity' and the term 'immaterial spirit'. Thus, he tells us that '[t]he "self" makes real physical identity' (NE: 237), and also that, when we would say that there is 'the same 'self' or individual', this is 'because of the same immaterial spirit' (NE: 235). Moreover, Leibniz also explicitly identifies the referents of 'individual' and 'soul': 'What makes the same human individual is...the soul' (NE: 241). All of this indicates that, at least in some circumstances, Leibniz uses 'I' to refer to one's own soul or spirit, as he also does, for example, in a passage from L 638 quoted above, where he identifies 'what is called 'I [*Moy*]' with "substance", "soul", "mind"'.

But now consider the expansion of the NE 241 passage just quoted:

What makes the same human individual is not 'a parcel of matter' which passes from one body to another, *nor is it what we call I*; rather, it is the soul. (NE: 241; our emphasis)

The overall context is one in which Leibniz has been at pains to distinguish real identity (which depends on identity of substance) from moral identity (which depends on identity of person or moral agent). And here Leibniz is explaining that sameness of individual is a matter of sameness of soul rather than sameness of matter or sameness of person. Thus, Leibniz is presupposing that *what we call I* is a *person*, and hence that the word 'I' can also be used to refer to a person. This, we might say on his behalf, is the *moral I* rather than the *physical I*.

How, then, does the ambiguity of the word 'I' help with the proper interpretation of (1) and (2)? When Leibniz commits to the proposition that *I am a particular immaterial substance*, he is clearly using the word 'I' in its first, meta-physical sense, rather than in the second, moral sense. What he is saying is *not* that I, *this particular moral person*, am a particular immaterial substance, for this would be obviously false. What he is saying instead is that I, namely, *this soul*

right here, is a particular immaterial substance, which is obviously true. But when Leibniz commits to the proposition that *it is metaphysically possible that I continue as an identical self-consciousness and identical self, independently of this particular substance*, he may reasonably be read as using the word ‘I’ in its second, moral sense, rather than in the first, metaphysical sense. Instead of saying that it is metaphysically possible that *this particular immaterial substance* continues as an identical self-consciousness and identical self, independently of this particular substance—which would be false if counterparts weren’t used to assess the truth-values of modal statements—Leibniz might simply be saying (truly, in his view) that it is metaphysically possible for *this moral person right here* to continue as an identical self-consciousness and identical self, independently of this particular substance. The two propositions, (1) and (2), properly read and understood, are mutually compatible, precisely because the word ‘I’ refers to an immaterial substance in the first and to a moral person in the second.²¹ On these readings, (1) and (2), taken together, are logically compatible with (3), i.e., the necessity of identity statements.

It might be suggested in reply that in the passages that support ascribing (2) to Leibniz, the word ‘I’ is used to refer to the soul rather than to the person. Interestingly, though, the word ‘I’ does not appear in these passages (except as referring to Leibniz himself when he speaks in the first person):

- (a) You [i.e., Locke] seem to hold, sir, that this apparent identity could be preserved in the absence of any real identity. Perhaps that could happen through God’s absolute power ... If a man could be a mere machine and still possess consciousness, I would have to agree with you, sir; but I hold that that state of affairs is not possible—at least not naturally. (NE: 236)
- (b) I admit that if God brought it about that consciousnesses were transferred to other souls, the latter would have to be treated according to moral notions as though they were the same. But this would disrupt the order of things for no reason ... (NE: 242)
- (c) I acknowledge that if all the appearances of one mind were transferred to another, or if God brought about an exchange between two minds by giving to one the visible body of the other and its appearances and states of consciousness, then personal identity would not be tied to the

21. It might be suggested that Leibniz distinguishes between *I* and *what we call ‘I’*. Perhaps, it might be said, Leibniz takes ‘I’ to refer to the soul and ‘*what we call “I”*’ to refer to the person. But the passage from L 638 indicates that this cannot be right, for in that passage Leibniz clearly uses the phrase ‘what is called “I”’—which is not relevantly semantically distinct from the phrase ‘*what we call “I”*’—to refer to the soul.

- identity of substance but rather would go with the constant appearances, which are what human morality must give heed to. (NE: 244)
- (d) Even if God were to change the real identity in some extraordinary manner, the personal identity would remain, provided that the man preserved the appearances of identity. (NE: 237)

Leibniz here supposes that it is logically possible, even if contrary to the order of nature, for the same consciousness, which defines personal (or apparent) identity, to be attached to different souls or minds (that are not 'really' or 'physically' identical). But it does not follow from this supposition that it is possible for me, *the soul that I am*, to continue independently of *this particular substance*. All that follows is that it is possible for me, the *person* that I am, to continue independently of *this particular substance*. And this, as we have seen, does not contradict the statement that I, the *soul* that I am, am *this particular substance*.²²

5. Conclusion

The two propositions that Wilson (1999) attributes to Leibniz and that she alleges to be inconsistent are of the following more general form:

- (1) I am numerically identical to X.
- (2) It is metaphysically possible that I continue to exist independently of X. (Thus, for example, it is metaphysically possible that I continue to exist even though X does not.)

These propositions appear to entail a contradiction: non-descriptive identity statements, such as (1), are necessary, and yet (2) appears to say that it is possible for (1) to be false. Past attempts to solve the problem denied that (1) is necessary, denied that Leibniz accepts (1), or denied that Leibniz endorses (2). We

22. An anonymous referee worries that the ambiguity solution assimilates Leibniz's view to Locke's view too much. On that solution, Leibniz and Locke would be disagreeing verbally but not metaphysically; yet Locke and Leibniz surely disagree on metaphysical issues, not merely linguistic ones. To this worry we offer the following response. As the referee recognizes, Locke would not accept the ambiguity of 'I'. This matters because it means that Locke, unlike Leibniz as he is understood on the ambiguity solution, denies that there is any reading of (1) on which (1) is true. For Locke, (1) is false, period. Metaphysically, Locke holds that the thinking thing in us might, for all we know, be material (E IV.3.6), whereas this is something Leibniz denies (NE: 379). Locke also holds that the person, or self, has parts (E II.27.17–18) and arguably that it is a substance (there is controversy about this, but see, e.g., Gordon-Roth 2015; Rickless 2015), whereas, on the ambiguity solution, the person (or referent of 'I'), for Leibniz, is either a partless substance or not a substance at all (for no true substance has parts).

have tried to explain why these attempts (proposed by Jolley, Vailati, and Bobro, respectively) fail. All of these attempts presuppose both that (1) and (2) are about the same thing (the sole referent of 'I') and that (1) and (2) are unambiguous. The modal solution denies the first presupposition, while the ambiguity solution denies the second. The modal solution identifies the fact that makes (2) true as a fact about counterparts of me and X, namely, that there is a possible situation in which counterpart-I (I^*) exists independently of counterpart-X (X^*), which is consistent with the fact that I am numerically identical to X. The ambiguity solution holds that 'I' is ambiguous, as between referring to a soul and a person, and that there is strong evidence that Leibniz explicitly endorses (1) only when 'I' refers to a soul and explicitly endorses (2) only when 'I' refers to a person. In that case, Leibniz would accept something of the form 'Y is numerically identical to X', while also accepting something of the form 'it is metaphysically possible for Z to continue to exist independently of X'. And, in doing so, Leibniz would not be contradicting himself. As far as we can tell, both the modal solution and the ambiguity solution are consistent with Leibniz's overall philosophical commitments, and it is not clear which of the two he would identify as his preferred solution to Wilson's conundrum on reflection.²³

Competing Interests

The authors have no competing interests to declare.

References

A. Works by Leibniz

- A *Sämtliche Schriften und Briefe*. Edited by Deutsche Akademie der Wissenschaften. Darmstadt, Leipzig, Berlin: Akademie Verlag, 1923–. Cited by series, volume, and page.
- AG *Philosophical Essays*. Edited and translated by Roger Ariew and Daniel Garber. Indianapolis: Hackett, 1989. Cited by page.
- CP *Confessio Philosophi: Papers Concerning the Problem of Evil, 1671–1678*. Edited and translated by Robert C. Sleight, Jr. New Haven and London: Yale University Press, 2005. Cited by page.

23. This article originated in an independent study on Leibniz's *New Essays* at UC San Diego in spring 2022. The modal solution was Min's idea; the ambiguity solution was Sam's idea. We are deeply grateful to Don Rutherford, the editors, and two anonymous referees for their constructive, thoughtful, and generous comments. Many thanks too to the copy editor, Anna Fare, for improving the article's readability.

- DM *Discours de métaphysique* [Discourse on Metaphysics]. In AG 35–68. Cited by section.
- G *Die Philosophischen Schriften von Gottfried Wilhelm Leibniz*. Edited by C. I. Gerhardt. Berlin: Weidmann, 1875–90. Cited by volume and page.
- L *Philosophical Papers and Letters*. Edited and translated by L. E. Loemker. 2nd edition. Dordrecht: Kluwer Academic Publishers, 1989. Cited by page.
- LA *The Leibniz–Arnauld Correspondence*. Translated and edited by H. T. Mason. Manchester: Manchester University Press, 1967. Cited by page.
- LGR *Leibniz on God and Religion*. Edited and translated by Lloyd Strickland. London and New York: Bloomsbury, 2016. Cited by page.
- Mon *Monadology*. In AG 213–25. Cited by section.
- MP *Philosophical Writings*. Edited and translated by Mary Morris and G. H. R. Parkinson. London and Melbourne: Dent, 1973. Cited by page.
- NE *Nouveaux Essais* [New Essays on Human Understanding]. Translated by Peter Remnant and Jonathan Bennett. Cambridge: Cambridge University Press, 1996. Cited by page numbers from A VI.6.
- T *Théodicée* [Theodicy]. In G VI. Cited by section number. Translated by E. M. Huggard. La Salle, Ill.: Open Court, 1985. Abbreviated [Hug] and cited by page.

B. Other Primary Sources and Secondary Literature

- Adams, Robert M. 1994. *Leibniz: Determinist, Theist, Idealist*. New York: Oxford University Press.
- Bobro, Marc E. 2004. *Self and Substance in Leibniz*. Dordrecht: Kluwer Academic Publishers.
- Descartes, René. 1984. *The Philosophical Writings of Descartes*. Translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch. Cambridge: Cambridge University Press. Cited by volume and page number of the Adam and Tannery edition. [AT]
- Gordon-Roth, Jessica. 2015. "Locke on the Ontology of Persons." *Southern Journal of Philosophy* 53 (1): 97–123. doi:10.1111/sjp.12098.
- Griffin, Michael V. 1999. "Leibniz on God's Knowledge of Counterfactuals." *The Philosophical Review* 108 (3): 317–34. doi:10.2307/2998464.
- Gut, Przemysław. 2017. "Leibniz: Personal Identity and Sameness of Substance." *Roczniki Filozoficzne* 65 (2): 93–110. doi:10.18290/rf.2017.65.2-5.
- Ishiguro, Hidé. 1972. *Leibniz's Philosophy of Logic and Language*. New York: Cornell University Press.
- Jolley, Nicholas. 1990. *The Light of the Soul: Theories of Ideas in Leibniz, Malebranche, and Descartes*. Oxford: Clarendon Press.
- Jorgensen, Larry M. 2019. *Leibniz's Naturalized Philosophy of Mind*. New York: Oxford University Press.
- Lewis, David. 1968. "Counterpart Theory of Quantified Modal Logic." *Journal of Philosophy* 65: 113–26.
- Locke, John. 1689/1975. *An Essay concerning Human Understanding*. Edited by Peter H. Nidditch. New York: Oxford University Press. Cited by book, chapter, and section number. [E]
- Mates, Benson. 1986. *The Philosophy of Leibniz: Metaphysics and Language*. New York: Oxford University Press.

- Mondadori, Fabrizio. 1973. "Reference, Essentialism, and Modality in Leibniz's Metaphysics." *Studia Leibnitiana* 5 (1): 74–101.
- Rickless, Samuel C. 2015. "Are Locke's Persons Modes or Substances?" In *Locke and Leibniz on Substances*, edited by Paul Lodge and Tom Stoneham, 110–27. New York: Routledge.
- Scheffler, Samuel. 1976. "Leibniz on Personal Identity and Moral Personality." *Studia Leibnitiana* 8 (2): 219–40.
- Vailati, Ezio. 1985. "Leibniz's Theory of Personal Identity in the *New Essays*." *Studia Leibnitiana* 17 (1): 36–43.
- Wilson, Margaret. 1976. "Leibniz: Self-Consciousness and Immortality in the Paris Notes and After." *Archiv für Geschichte der Philosophie* 58 (4): 335–52. Reprinted in *Ideas and Mechanism: Essays on Early Modern Philosophy*, 373–87. Princeton: Princeton University Press, 1999.
- Wilson, Margaret. 1979. "Possible Gods." *The Review of Metaphysics* 32 (4): 717–33. Reprinted in *Ideas and Mechanism: Essays on Early Modern Philosophy*, 407–20. Princeton: Princeton University Press, 1999.

Submitted: 15 July 2024 **Accepted:** 06 May 2025 **Published:** 17 November 2025

