



HAL
open science

Identity-less personal past: the dependence of the neo-lockean theory of personal identity on the causal theories of memory

Anita Lee Vanbeveren

► To cite this version:

Anita Lee Vanbeveren. Identity-less personal past: the dependence of the neo-lockean theory of personal identity on the causal theories of memory. *Philosophy*. 2025. <dumas-05535821>

HAL Id: dumas-05535821

<https://dumas.ccsd.cnrs.fr/dumas-05535821v1>

Submitted on 4 Mar 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License



Anita VANBEVEREN-LEE

Identity-Less Personal Past
The Dependence of the Neo-Lockean Theory of Personal
Identity on the Causal Theories of Memory

Mémoire de Master 2 « Sciences humaines et sociales »

Mention : Philosophie

Parcours : Philosophie de la cognition

Sous la direction de M. Denis PERRIN

Année universitaire 2022-2024



Anita VANBEVEREN-LEE

Identity-Less Personal Past
The Dependence of the Neo-Lockean Theory of Personal
Identity on the Causal Theories of Memory

Mémoire de Master 2 « Sciences humaines et sociales »

Mention : Philosophie

Parcours : Philosophie de la cognition

Sous la direction de M. Denis PERRIN

Année universitaire 2022-2024

To my uncle, Xian-Hu Lee, who first
introduced me to the joys of critical thinking.

Acknowledgements

Writing this thesis has been one of the most challenging yet rewarding projects of my life. When I first began, I was unfamiliar with analytic philosophy and philosophy of mind, but thanks to the invaluable support and encouragement of many people, I was able to produce a finished and meaningful work despite numerous difficulties.

First and foremost, I would like to express my deepest gratitude to my thesis director, Professor Denis Perrin, for his patient guidance, insightful feedback, and kindness in addressing my sometimes stubborn naivety on complex philosophical points. Without his constant support, this thesis would not have been possible.

I am also grateful to Professors Kirk Michaelian and Rémi Clot-Goudard, whose inspiring lectures and thoughtful responses to my questions greatly enriched this research. Their professionalism and generosity in bearing with my naïve inquiries have been a true encouragement.

My peers have also been a source of strength throughout this journey. Their willingness to engage in discussions and share their company helped me persevere through this demanding process that few around me fully understand.

Finally, I wish to thank my family for their unwavering patience and support – both emotional and financial – during the many difficult moments over these past three years. Without them, I would not have made it through this marathon.

Table of Contents

DECLARATION SUR L'HONNEUR DE NON-PLAGIAT	3
ACKNOWLEDGEMENTS	5
TABLE OF CONTENTS	6
0 INTRODUCTION	8
0.1 THE LOCKEAN IDENTITY	10
0.2 BUTLER AND REID, AND THE CIRCULARITY OBJECTION	14
0.3 QUASI-MEMORY AND MEMORY THEORIES	18
0.4 OVERVIEW	21
PART I: FOUNDATIONS & THE PROBLEM	24
1 METAPHYSICS OF PERSISTENCE	25
1.1 PERSISTING AS PERDURING OR ENDURING?	26
1.2 TEMPORAL GROUNDING OF THE REGION	30
1.3 TEMPORAL GROUNDING OF THE OBJECT	32
1.4 CONCLUSION	34
2 PERSONAL IDENTITY THEORIES	37
2.1 MEMORY-BASED THEORIES	38
2.2 BODY-BASED THEORIES	41
2.3 HYBRID THEORIES	43
2.4 CONCLUSION	49
PART II: INADEQUATE SOLUTIONS & THEIR LIMITS	50
3 FROM CAUSAL TO POST-CAUSAL THEORIES OF MEMORY	51
3.1 CAUSAL THEORIES OF MEMORY	52
3.2 CAUSALISM AND CIRCULARITY	54
3.3 POST-CAUSAL THEORIES	57
3.4 POST-CAUSALISM AND CIRCULARITY?	60
3.5 CONCLUSION	64
4 THE FIRST PERSON, IEM, AND QUASI-MEMORY	65
4.1 'I'-THOUGHTS AND NON-OBSERVATIONAL KNOWLEDGE	66
4.2 MEMORY-BASED 'I'-THOUGHTS AND QUASI-MEMORY	69
4.3 IEM AND PERSONAL MEMORY	73
4.4 CONCLUSION	78
PART III: TOWARD A NEW FRAMEWORK	79
5 THE FEELING OF MINENESS AND REMEMBERING	80
5.1 SCHECHTMAN'S DILEMMA	82
5.2 R.B.'S MINENESS-LESS MEMORIES	86
5.3 FERNÁNDEZ ON THE FEELING OF OWNERSHIP	88
5.4 CONCLUSION	91
6 IDENTITY-LESS MEMORY-BASED PERSONAL IDENTITY	93

Table of Contents

6.1 A COMPATIBILITY TEST FOR MEMORY THEORIES.....	94
6.2 THE RELEVANCE OF EPISTEMIC RELEVANCE	98
6.3 A SIMULATIONIST MEMORY-BASED PERDURANTIST PERSONAL IDENTITY FRAMEWORK.....	103
6.4 CONCLUSION.....	108
CONCLUSION.....	108
BIBLIOGRAPHY	113

0 Introduction

It is uncontroversial that personal identity means a great deal to each of us, and it is similarly uncontroversial that our personal memory has something special to do with our personal identity (Copenhaver, 2018; Fernández, 2019; Gordon-Roth, 2023; Klein & Nichols, 2012; Locke, 2014; Michaelian, 2016; Noonan, 2019; Parfit, 1971, 1984; Perry, 2008; Piolino et al., 2006; Roache, 2016; Sattig, 2018; Schechtman, 1996, 2011a, 2024; Slors, 2001; Trakas, 2021; Williams, 1973a; and this list is non-exhaustive).

Being one of the most ancient philosophical questions, personal identity can be approached through several different lines of inquiry (Noonan, 2019; Olson, 2023; Perry, 2008; Schechtman, 1996), including the constitution, persistence, and reidentification questions, among others. Memory, on the other hand, has come to occupy an important place in the contemporary literature, especially with the flourishing of empirical research since the late 20th century (De Brigard, 2023; Michaelian, 2016). The self has been an important aspect of more recent memory studies, under the influence of the 18th-century English philosopher John Locke (Perry, 2008), who initiated the tradition of personal identity theories based on the ‘sameness of consciousness’, commonly interpreted as the faculty of memory: remembering allows us to entertain the sameness of consciousness.

Memory, as a natural kind, can be further divided into subcategories or kinds of memory, one of which appears to be particularly intriguing. Endel Tulving, the most prominent and frequently cited contributor to the psychological literature on memory, characterizes ‘episodic remembering’ as “to consciously re-experience past experiences” (Tulving, 2002). This has evolved into the notion of mentally ‘re-living’ a personal past event (Klein, 2015; Michaelian, 2016; Perrin et al., 2020; Suddendorf & Corballis, 2007; Tulving, 2002: 6). *Prima facie*, ‘re-living’ and ‘re-experiencing’ both sound mysterious, even magical, and they raise many questions about this special relation between the self and its memory. Among these questions, and stemming from the neo-Lockean view of personal identity, is the numerical

1 Metaphysics of Persistence

identity question, which faces the well-known circularity objection to the psychological continuity criterion.

As fiercely criticized by Butler and Reid, Locke's original version of the memory theory of personal identity appears too strong and rigid to escape the problem of circularity (Perry, 2008), wherein the definition of memory presupposes personal identity and therefore cannot constitute it. Neo-Lockean theorists have done much to defend the memory-based view from this objection. One of the most influential strategies has been the development of the notion of 'quasi-memory', which involves loosening and redefining the strict causal connection that binds memory and personal identity into a circular structure (Shoemaker, 2003; Parfit, 1971, 1984). If the causal condition of memory can be loosened, does this imply that memory causation comes in degrees or different kinds? If so, where exactly is the boundary between remembering and imagining, for instance? By reconsidering this point, might it be possible for the neo-Lockean theory to dispense with any form of memory causation – and thus no longer be subject to the circularity objection?

Among theories of memory, one appears particularly well-suited to addressing these questions: the theory that conceives of remembering as simulating the past – namely, the simulation theory. The simulation theory of memory offers an empirically grounded framework for distinguishing remembering from imagining, claiming that they are, in fact, mental states of the same kind (Michaelian, 2016). This claim faces obvious challenges – for example, how can this view account for the distinctiveness of memory, which provides us with reliable information about the world? Yet it is arguable that simulationism can accommodate this distinction through the framework of source monitoring, which ensures the relative reliability of memory in contrast to its siblings: pure imagination and counterfactual thinking.

The simulation view of memory is attractive for another reason: it does not seem to fall into the kind of circularity that afflicts memory in the Lockean theory of personal identity. If the circularity problem lies in the presupposition of personal identity within the definition of memory, then this condition must be removed for a neo-Lockean, memory-based theory of personal identity to be sound. What makes radical simulationism (Michaelian, 2024) especially promising in this regard is not just its rejection of the causal connection condition, but more crucially, its rejection of the 'previous experience condition' – that is, the requirement that the remembered event must have been personally experienced by the subject. Versions of simulationism that retain the previous experience condition still fall prey to the circularity

problem, since they presuppose the very identity they aim to explain. By dropping this condition, radical simulationism avoids such presuppositions and offers a non-circular foundation for memory-based accounts of personal identity.

The title of this thesis announces its objective: to find a definition of memory that is identity-less: that is, one that does not presuppose personal identity, so that memory can, in turn, constitute personal identity. Throughout this research, I will aim to demonstrate that such a theoretical position can be achieved by adopting and aligning with simulationism. This is not merely a matter of loosening the causal condition to arrive at a weaker version of quasi-memory; rather, it is a move that enables the neo-Lockean view of personal identity to avoid circularity while still defining personal identity through the sameness of consciousness, as originally proposed by Locke.

0.1 The Lockean Identity

To begin with, what exactly is the problem of personal identity? Do persons have a special kind of identity, or is it the same as that of other objects or animals? In *An Essay Concerning Human Understanding*, Locke addresses the topic of identity in Chapter 27, turning specifically to personal identity only in Section 9, after establishing a general account of the identity of things. To do so, he clarifies several fundamental notions that distinguish identity from diversity. Together, these notions form the basis of the Lockean account of personal identity.

First, the subject matter of Lockean identity is, in principle, *diachronic*: that is, when “a person P_2 at a time t_2 [is] the same person as person P_1 at an earlier time t_1 ” (Swinburne, 1976: 223; Noonan, 2019: 84). This notion contrasts with *synchronic* identity: that is, when “two persons at a given time [are] the same person” (Swinburne, 1976: 228; Noonan, 2019: 85). Diachronic identity is often characterized as the persistence question of personal identity: how does a person persist through time? As Locke writes: “when, considering anything as existing at any determined time and place, we compare it with itself existing at another time, and thereon form the ideas of identity and diversity” (Locke, 2008: 33).

Secondly, within the persistence question, we are more specifically concerned with *numerical* identity, which should be distinguished from other identity-related issues, such as

1 Metaphysics of Persistence

character identity (Schechtman, 1996) and *personhood* (Olson, 2003).¹ Numerical identity refers to how we can be sure “that it is that very thing, and not another, which at the same time exists in another place, how like and undistinguishable soever it may be in all other respects” (Locke, 2008: 33).² It is a concern with individuality across both space and time: a notion that applies not only to persons, but also to things and animals. For example, we might ask: How do we know that this table in front of me is the same one I saw in this position yesterday? Or: How do we know that the fish on my plate right now is the same one that was in my bucket yesterday? These are persistence questions concerning the numerical identity of an object and an animal.

Thirdly, considering the nature of persistence, Locke is an *endurantist*. That is, he holds that for a thing to have its complete identity, it must persist through time as a whole. This concept contrasts with that of the *perdurantist*, who considers objects as having temporal parts, the assembly of which constitutes the diachronic numerical identity of objects. Locke uses the examples of motion and thought to explain his endurantist concept of identity, describing them as “both which consist in a continued train of succession”, and since “each perishing the moment it begins, they cannot exist in different times, or in different places,” no motion or thought “can be the same, each part thereof having a different beginning of existence” (Locke, 2008: 34–35). This idea is based on his rule that one thing cannot have two beginnings, and two things cannot share one and the same end.

Lastly, Locke differentiates ‘intelligence’ and ‘body’ as substances of different kinds. Setting God’s mystical kind of substance aside, the other two sorts of substances “exclude any of the same kind out of the same place” (Locke, 2008: 34), since two bodies cannot occupy the same place at the same time, nor can one intelligence be in two places simultaneously.³ However, the changing body of a machine, an animal, or a human seems to be guaranteed by an overall identity reassured by “one common life” (Locke, 2008: 36). That is, something other than the composed body is responsible for the identity of the object. For example, I am

¹ The typical formulation of the characterization question is: “What are the special properties that make me the person I am?” The personhood question is typically formulated as: “What does it take to be a person?” The latter can also be understood as the *constitution* question.

² Although my primary focus is on diachronic identity, introducing the concept of numerical identity requires a brief mention of synchronic identity and diversity, since it is impossible “that two things of the same kind should exist in the same place at the same time” (Locke, 2008: 33).

³ Notice how it isn’t formulated the other way around: “two intelligences can’t be in the same place at the same time, and one body can’t be in two places at the same time.”

intuitively inclined to recognize a grand oak tree as the same as the tiny oak tree sprout I planted when I was little, provided I have grounds to believe that the entirety of the tree has never been removed or replaced – despite the fact that the particles composing the sprout and the mature tree are neither exactly the same in quantity nor in quality. This being said, it seems that as long as there is the “same organized body” (Locke, 2008: 36–37), which includes the participation of fleeting and joining particles, identity will be assured.⁴ Nonetheless, the matter becomes more complicated. For example, if we imagine a man’s soul in a hound’s body – a typical body-switching thought experiment – the hound’s body will have been the same continuous organized body since its birth, but it is unclear whether the resulting “man-hound” is the man or should be considered the hound.⁵ Highlighting the complexity of the identity question, Locke further distinguishes three kinds of identity: that of the same substance, the same man, and the same person (Locke, 2008: 37). The same substance corresponds to the same kind of thing, namely the substance of the “intelligent thinking thing” and that of the body; the same man refers to the body, or the materialistic concept of individualization; and the same person is a legal, forensic concept that ties the individual to related responsibilities and moral behaviors.⁶

Building on these premises, Locke provides his readers with a definition of ‘person’ as:

a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places; which it does only by that consciousness which is inseparable from thinking, and as it seems to me, essential to it: it being impossible for any one to perceive without perceiving that he does perceive.

(Locke, 2008: 39)

Therefore, according to Locke, it is the continuity of consciousness that constitutes personal identity. Personal identity can extend as far back (or forward) in time as one’s consciousness can reach, “so far reaches the identity of that person” (Locke, 2008: 39).⁷ Typically, the

⁴ As body-based theories and animalism would claim.

⁵ The original example of this is Heliogabalus’s soul taking over a hog’s body (Locke, 2008: 37).

⁶ It is possible to observe a similarity between Locke’s definitions of substance, body, and person, and the Christian concept of the Trinity: The Holy Spirit, Jesus Christ, and God. However, Locke’s taxonomy is not strictly followed in subsequent personal identity theories. In particular, the term ‘person’ is often applied to quite different concepts. It is important to understand Locke’s original intention in naming ‘person’, how it differs from other kinds of identity, and to recognize the influence of his religious context.

⁷ In his text, Locke doesn’t explicitly mention future projection, aside from a brief comment on the indeterminate duration of consciousness and the possibility that one’s self “may be the same self by the same consciousness continued on for the future” (Locke, 2008: 50). However, in later chapters, I will argue that future-projecting self-

1 Metaphysics of Persistence

consciousness that reaches back to our personal past experiences and actions will later in the text be identified as personal memory, more specifically classified as the experiential memory Tulving calls episodic memory. Compared to either material or immaterial substances, Locke claims that consciousness alone suffices and is necessary for personal identity (Locke, 2008: 39–45). Personal identity consists “not in the identity of substance, but [...] in the identity of consciousness” (Locke, 2008: 46). That is, personal identity consists not in the sameness of either the “intelligent thinking thing” or the body, but in the unceasing continuity of consciousness.

Locke himself talks about consciousness in a broad and general sense, leaving room for disagreement among his interlocutors. Although he does not explicitly state in his texts that memory is what he means by consciousness, there are quite obvious – though implicit – suggestions that memory is that consciousness, since memory is what connects past and present consciousness. According to Tulving (1972), consciousness can be further categorized into three different levels of awareness: unconsciousness or ‘anoesis’, consciousness or ‘noesis’, and self-consciousness or ‘autonoesis’. Each of these three levels of consciousness can be paired with a kind of memory. Anoesis corresponds to the kind of consciousness involved in non-declarative memory, including procedural memory; noesis corresponds to semantic memory about facts; and autonoesis corresponds to episodic memory, which, as mentioned earlier, is the special kind of memory that supposedly allows one to remember only one’s own past experiences. This last kind of consciousness seems to be the most compatible with what Locke had in mind.

A number of objections have surfaced since the publication of Locke’s *Essay*. The first concerns the prominence of taking memory rather than the body as the basis for personal identity. Shoemaker, for example, has argued that although the memory criterion is fundamental, it is necessary to consider the body criterion to verify the correctness of memory: except in exceptional cases where the latter might not be required (Shoemaker, 2008). I find this view both attractive and favorable to the hypothesis of the present research: despite the primacy of consciousness and memory for personal identity, I would like to keep open the possibility, in Shoemaker’s sense, that the body criterion might be a necessary condition in most common cases in our world, although its absence may be acceptable under certain

consciousness plays the exact same role as past-remiscing self-consciousness in mental time travel and in constituting personal identity.

circumstances (Shoemaker, 2008). If we consider Shoemaker a hybrid theorist (Noonan, 2019: 201) – that is, a neo-Lockean open to the necessity of some bodily condition: there are more radical body theorists who regard the body criterion as the sole necessary and sufficient condition for personal identity, such as Olson (2003) and Snowdon (2014) (Noonan, 2019: 201–202). While I find the hybrid approach comprehensive – adding the body criterion to strengthen the correct use of memory in defining personal identity – I find it quite implausible that personal identity consists solely in body, without any condition of continuity of mind.

Among those who agree that personal identity consists (at least in part) in the mind rather than (solely) in the body, there are different views on the psychological aspects that constitute personal identity (Reid, 2008a, 2008b; Parfit, 1984). It is true that the mental life of an individual involves not only memory but also many other psychological properties such as intentions, beliefs, desires, emotions, and personality traits. For this reason, some philosophers support a broader claim for psychological connection without placing memory in a privileged position (Parfit, 1984). However, I tend to agree with Shoemaker in considering memory the fundamental pillar that enables the other psychological connections (Shoemaker, 2003). It seems not only plausible but also intuitive to suggest that without memory playing its role, there would be no way for intentions, beliefs, desires, emotions, and personality traits to occur and make sense. That is, I suggest that although personal identity consists in the continuity of the overall psychological environment, memory plays a fundamental role without which this environment cannot exist. Thus, in this work, I will use ‘mental-’, ‘memory-’, and ‘psychology-based’ interchangeably when describing accounts, noting minor differences only when necessary. Proponents of this memory-based view of personal identity form the ‘neo-Lockean’ school of personal identity – a series of arguments aligned with Locke’s principal ideas.

Let us now turn to the most problematic objection against the original Lockean theory, which, unfortunately, is inherited by the neo-Lockeans: the problem of circularity.

0.2 Butler and Reid, and the Circularity Objection

Locke’s view that memory is the sole necessary and sufficient criterion for personal identity has been widely criticized as too strong and overly strict. Since then, neo-Lockean theorists have sought ways to relax this condition to a reasonable degree, allowing the theory to accommodate common human experiences such as forgetting and periods of sleep. In doing

1 Metaphysics of Persistence

so, neo-Lockeans also address the circularity objection raised by Butler and Reid, aiming to defend a memory-based theory of personal identity. Before examining the proposed solutions, it is important to first clarify the nature of their objection to Locke.

Butler finds it obvious why memory in terms of consciousness cannot define personal identity, stating that “one should really think it self-evident, that consciousness of personal identity presupposes, and therefore cannot constitute, personal identity, any more than knowledge, in any other case, can constitute truth, which it presupposes” (Butler, 2008: 100). He also claims that “though consciousness of what is past does [...] ascertain our personal identity to ourselves”, it is absurd to say that it is necessary for one being the same person, because that would imply that one is never oneself when not conscious: in cases of forgetfulness, moments asleep, or drunkenness. This is perhaps best illustrated by the three objections Butler provides.

First, based on his analysis that a person is not continuously conscious throughout life, Butler considers Locke’s thesis to make personal identity a transient notion, which makes no intuitive sense: in daily life, we do not think of ourselves as one person yesterday, another today, and yet another tomorrow, simply because we sleep each night. Secondly, Butler distinguishes two different ways of speaking of sameness: either in a “loose and popular sense”, or in a “strict and philosophical” manner (Butler, 2008). He criticizes Locke’s use of personal identity as adopting the former sense, whereas Butler himself thinks in the latter. Accordingly, he differentiates between the ‘truth of the personal past’, i.e., one’s successive actions including enjoyments and sufferings, and one’s ‘continuous consciousness’, i.e., remembering and forgetting. According to Butler, the former notion precedes the latter, because it adheres to a strict and philosophical sense of identity, whereas the latter corresponds to the loose and popular sense (Butler, 2008: 103–104). Thirdly, Butler argues that the person is not consciousness itself, but one can be conscious of being the same substance or the same property of a substance. That is, Butler considers persons as substances or properties of substances.

Following Butler’s lead, Reid outlines similar objections. Basing his analysis of identity on Leibniz’s law of the indiscernibility of identicals, Reid claims that memory, like other mental operations, is not where the identity of a person resides, because “the identity of a person is *perfect* identity: wherever it is real, it admits of no degrees; and it is impossible that a person should be in part the same, and in part different; because a person is a monad, and is not

divisible into parts” (Reid, 2008a: 111, emphasis added). In the absence of contrary evidence, the mind tends to recognize similarity between oneself in a piece of memory and oneself at present reminiscing that memory. According to Reid, this suffices as evidence of a person’s “perfect identity”. Similar to Butler’s distinction between a “strict and philosophical” sense and a “loose and popular” sense of identity, Reid distinguishes the perfect identity of a person, which is indivisible and indiscernible, from the imperfect identity ascribed to bodies, where the term ‘identity’ is used merely for convenience of speech (Reid, 2008a: 112).

Several parallels between Butler and Reid’s arguments can be observed. First, both accuse Locke’s concept of identity of being “loose and popular”, or ill-applied merely for convenience of speech, whereas their own concepts of identity are “strict and philosophical” or “perfect”. The “loose and popular” sense of identity, as Butler puts it, corresponds to what Reid identifies as mere similitude and sameness in kinds and species, as opposed to sameness of individuation. The basis of their critique lies in their shared interpretation that something as constant as personal identity individuation should not be based on consciousness – which, according to them, suffers discontinuities. This is why Locke’s theory appears too rigid: it cannot accommodate the most common cases of discontinued consciousness.

Answering the problem of the over-strictness of requiring memory continuity as mandatory, Grice (1941) proposed a method that secures a less direct kind of continuity, allowing one to avoid the Brave Officer Paradox illustrated by Reid (2008b) and other similar counterexamples to the original memory theory involving forgetfulness and momentary discontinuities of the mind. In the Brave Officer case, the older general no longer needs to remember himself as a boy robbing an orchard to justify his identity as both; it suffices that the older general remembers himself taking a standard from the enemy in his first campaign as a young man, who in turn still remembers robbing the orchard as a boy. This means that between times t_1 , t_2 , and t_3 , there is not complete continuity but a sequential, chained continuity: although t_1 and t_3 are not directly connected, they share continuity enabled by t_2 ’s continuity with both t_1 and t_3 . These temporary moments that form a chain of continuity are called ‘Total Temporary States’ (TTSs). Grice’s TTSs fairly loosen Locke’s original strict condition of continuity, while still preserving the principle of continued consciousness. However, unfortunately, Grice’s TTS structure does not escape the other troublesome objection: circularity (Perry, 2008).

1 Metaphysics of Persistence

Both Butler and Reid also point out that memory presupposes personal identity and therefore cannot constitute it. Even if one stream of consciousness can be broken down into discontinuous parts that compose a continued chain, it seems that each of those discontinuous parts still presupposes personal identity, which is precisely the element that chains them together. The only way to distinguish the veridicality of a memory is through personal identity: I know I remember correctly that I ate an apple yesterday because I am the person who ate that apple yesterday. I can confidently enunciate this knowledge because I actually remember the feeling of holding that apple in my hand and taking a bite. It is because I can remember that the person who went to bed yesterday was myself, that I can know – albeit despite the temporary discontinuity – that the person waking up from the same bed right now, gaining present consciousness, is the same person who went to bed last night. The fact that I am the same person is, in turn, the only reason why I can remember going to bed the night before. It seems that this chaining element of personal identity plays a crucial role as evidence of memory causation.

The main reason contemporary memory philosophers and psychologists are particularly interested in episodic memory is that it seems to be the kind of memory that not only implies the self, but also provides a direct access to a past state of the world. The key difference that distinguishes episodic memory from other kinds of memory and other mental states, such as imagination and counterfactual thought, was pointed out by Martin and Deutscher (1966): episodic memory is special because of the *causal connection* it offers, which accounts for the continuity of our personal existence. According to the causal theory of memory, there are basically three conditions for remembering (Martin & Deutscher, 1966). One can say they remember a past event if (1) they can represent the event at the time of retrieval, (2) they represented the event at the stage of encoding, and (3) there is an appropriate causal connection between the representation at retrieval and at encoding (Michaelian & Robins, 2018).

Although all three conditions need to be met for one to remember, condition (3) is of particular interest in the context of this research, as it seems to be responsible for the circularity problem in memory-based personal identity. The appropriate causal connection means that the remembering individual must recall the event without any external assistance, such as testimony. This causal condition of episodic memory is a crucial reason why neo-

Lockean theory remains prone to circularity. We can articulate the circular neo-Lockean theory as follows:

Personal identity over time depends on a causal memory connection between person *P1* at time *t1* and person *P2* at time *t2*. Person *P2* identifies as *P1* because she remembers being *P1* at *t1*.

Following this definition, the circularity objection can be stated similarly:

Person *P2* identifies as *P1* precisely because she remembers being *P1*. But this memory already assumes personal identity to verify its accuracy. Therefore, defining personal identity through memory is circular.

As Reid's dilemma famously puts it, Locke would either have to admit that one and the same person could have two beginnings of memory or that his memory theory is viciously circular (Perry, 2008; De Brigard, 2023: 33). In other words, if Locke wishes to define personal identity through memory without circularity, he must accept that a person can have multiple memory beginnings allowing for forgetting and temporary unconsciousness. However, it is important to clarify two points. First, Locke's original account of personal identity through memory does not explicitly require a causal connection as part of the definition of memory itself. Second, the more modern causal theory of memory introduces its own circularity problem: it requires that remembering involves a causal link to one's past experiences, which presupposes the persistence of personal identity that memory is supposed to explain.

0.3 Quasi-Memory and Memory Theories

One prominent attempt to save the neo-Lockean view is to replace the concept of ordinary memory with that of 'quasi-memory' (Shoemaker, 2003; Parfit, 1984), a form of remembering that loosens the usual causal condition found in typical memory. When someone quasi-remembers an event, the event in question may have been experienced by someone other than the quasi-rememberer. This means that people can remember experientially – by virtue of episodic remembering – events they did not themselves experience. In this case, personal identity is no longer embedded within (quasi-)memory, allowing it to define personal identity without circularity.

Even though quasi-memory offers an attractive loosening of the causal theory of memory that could ground personal identity over time, contemporary proponents of the circularity objection, such as Schechtman (1990), remain seriously skeptical. She argues that a quasi-memory candidate would necessarily either be a delusion or presuppose personal

1 Metaphysics of Persistence

identity if genuine. The fundamental reason for this, she explains, lies in the feeling of ‘mineness’ – the sense of ownership – attached to the episodes of a person’s past experiences; this cannot be separated from the content of the memory. Thus, “to define a state that can be properly called one of the states constitutive of personal identity, we will have to presuppose the existence of a persistent person who is the subject of that state” (Schechtman, 1990: 87). The requirement that memory must involve a feeling of ‘mineness’ – that is, a subjective sense of ownership over the remembered event – has often been taken as a phenomenological marker of genuine episodic remembering. Some theorists treat this condition as conceptually independent of any particular metaphysical theory of memory, and it need not entail a commitment to causalism. However, when combined with the causal condition, the mineness criterion can appear to reinforce the idea that remembering must be tied to one’s own past experiences, thus potentially contributing to the circularity problem for memory-based accounts of personal identity.

Recent debates in the phenomenology of memory shed light on how this mineness condition might be more flexibly understood. For instance, Klein and Nichols (2012), discussing the case of patient R.B., suggest that the phenomenological feeling of ownership is not a necessary feature of episodic memory, as R.B. seems to remember past events without experiencing them as his own. Fernández (2019), however, disputes this interpretation, maintaining that R.B.’s case does not demonstrate a breakdown in the sense of personal identity, and that the memory is preserved despite the attenuated phenomenology. This disagreement reopens the broader question: what are the necessary and sufficient conditions for an episodic memory to qualify as such, and how do these conditions contribute to the constitution of personal identity?

Among personal identity theories, one has completely abandoned the psychological continuity condition: ‘strong animalism’, or what Noonan (2019: 201) calls ‘biological animalism’, mainly represented by Olson and Snowdon (Noonan, 2019: 201–202; Snowdon, 2014; Olson, 1997, 2003). This position adopts a biological approach that contrasts with psychological perspectives based on consciousness or memory. Proponents of strong animalism present compelling arguments that have influenced recent developments in psychological continuity theory. A key argument concerns counting fetuses as persons: since organisms are animals and persons are animals, living fetuses, as organisms, should also be classified as human animals and, therefore, persons. This poses a challenge for neo-Lockean

theorists, who struggle to determine the precise moment when a fetus or young child becomes a legitimate person. Is there a specific degree of consciousness required for entry into the class of persons?⁸ What role do memory and consciousness play in addressing this dilemma? These are crucial questions that neo-Lockean proponents must confront. In contrast, because the human animal body requires no additional condition for numerical identification, strong animalism effectively avoids the circularity problem faced by psychological continuity theory.

On the one hand, neo-Lockean theorists need to address the newly reinforced circularity objection through a more detailed investigation of the phenomenology of remembering and its role in episodic memory. On the other hand, it is desirable for them to develop a strategy that avoids falling under strong animalism by providing an account that requires a certain form of sameness of consciousness. It is crucial to note that animalism does not address the same questions about personal identity as memory-based theories. The former more straightforwardly answers the question, “What are we?”, while the latter addresses, “What does it take for a person to persist through time?” (Noonan, 2019: 199). Neo-Lockean theories may not fully answer the ontological question of what a person is, but they do provide conditions for qualifying person re-identification. Therefore, the strategy to confront strong animalism would be to argue that psychological continuity theory cannot be entirely dismissed, as it is at least partially necessary for re-identifying a person, even if it is not sufficient on its own.

This causal connection, which according to Martin and Deutscher (1966) is a crucial distinction between memory and other mental states, has recently been challenged by the simulation theory of memory. The simulation theory shifts the perspective from a unidirectional retrieval of information to a more symmetrical process that allows mental exploration of both the past and the future. This debate is also referred to as the (dis)continuism debate (Perrin, 2016; Perrin & Michaelian, 2017; Sant’Anna, 2021).⁹ This symmetry, which allows us to mentally go back and forth in time at will, is called ‘mental time travel’ (MTT): it connects our past and future into a continuity.

⁸ This will not only help distinguish persons from fetuses, but also from other conscious animals, especially those that are very similar to us, such as certain primates.

⁹ Proponents of discontinuism argue that the mechanisms of remembering and imagining are fundamentally *asymmetric*. Conversely, continuism advocates maintain that these mechanisms are *symmetrical*.

1 Metaphysics of Persistence

Beyond its intriguing name, MTT has sparked interesting discussions about what this concept reveals about ourselves. If a person is not confined to simply remembering the past or imagining the future in the traditional sense, but can also ‘remember’ the future (in the sense of having a certain nostalgic feeling that traditionally belongs uniquely to remembering the past, or remembering ‘having to do something’) and ‘imagine’ the past (in the sense of asking the question: ‘What would have happened if I had done this instead of that?’), what implications does this have for our understanding of personal identity? When we lift these confinements, what does it suggest about the causation involved in the remembering process? Should we conclude that causation does not exist at all? Alternatively, could it exist in both directions of mental information generation and retrieval, but in a weaker sense or to a lesser degree? How does this shift in research trajectory affect the neo-Lockean theory of personal identity? Lastly, what defines psychological continuity when stripped of its causal condition?

Even though there are quite a few rival theories of memory, not many differ radically from the causal theory regarding condition (3), the requirement of a causal connection. If this causal theory of memory is incorrect, however, it would mean that this kind of causal connection isn’t necessary for an individual to remember. If so, memory would no longer presuppose personal identity and could define it without constraint. In other words, if there is a viable alternative to the causal theory of memory, there might be a chance for neo-Lockean personal identity theory to avoid the circularity objection by redefining memory and remembering.

This brief overview highlights the challenging questions raised by recent research on personal identity and memory: two fields that, I realize, are increasingly developing independently. These observations motivate the core project of this thesis. The ambition here is to explore possible answers by bringing these topics back into dialogue. Hopefully, this thesis will draw attention to the potential of the simulation theory of memory to support a version of neo-Lockean theory that is not only non-circular but also offers a better explanation of the relationship between memory and personal identity.

I will now proceed with an outline of the thesis.

0.4 Overview

Here is an overview of what will roughly follow this introductory chapter. Each chapter can be considered a premise that builds up to my general working hypothesis: that the simulation

theory of memory is better than causal theories of memory in supporting a non-circular, memory-based neo-Lockean theory of personal identity.

Part I aims to lay the foundation of this research and highlight the central problem. Chapter 1 illustrates the ontological problem of identity over time through a discussion of the metaphysics of persistence. The objective is to establish the metaphysical framework necessary to understand diachronic personal identity. Particular attention is given to the perdurantism/endurantism debate, which sets out the ontological assumptions about what kinds of entities persist over time and which persistence structure best supports psychological continuity. Chapter 2 provides a detailed overview of various accounts of personal identity, aiming to assess whether a reconciliation between memory-based and body-based theories is possible. It ultimately favors a hybrid version of the neo-Lockean account, similar to Shoemaker's general position.

Part II demonstrates the inadequacy of traditional solutions to the circularity problem. Chapter 3 discusses Martin and Deutscher's causal theory of memory within a broader review of memory theories developed over recent decades. Causal theories are shown to be inadequate for grounding any memory-based account of personal identity unless the causal condition is revised, as they fail to account for the subjective ownership of memory without presupposing identity. In Chapter 4, the self as a linguistic object is analyzed through the lens of Anscombe's work. Drawing on her views on agency and action, I argue that the use of 'I' does not designate any object that is the self. Applied to memory claims, this suggests an instability in personal identity that is not easily grasped intuitively. However, this point can be illustrated through possible worlds where quasi-remembering is both possible and commonplace.

The final part of this thesis, Part III, introduces a potential new framework for a memory-based theory of personal identity that does not presuppose identity. To address the renewed circularity objection raised by Schechtman, Chapter 5 offers a detailed discussion of the phenomenology of remembering. It is argued that the feeling of mineness is not dependent on a causal connection between the rememberer and the original event. Moreover, this feeling is contingent in such a way that it cannot serve as a reliable criterion for memory. In Chapter 6, theories of memory are analyzed and compared in light of the epistemic relevance condition, identifying the simulation theory as the most suitable candidate to support a non-circular, identity-less, memory-based neo-Lockean theory of personal identity.

1 Metaphysics of Persistence

The final section presents a personal identity framework that adopts a perdurantist view of persistence, integrates a hybrid model with bodily grounding, and emphasizes the necessity of simulational remembering.

A concluding chapter follows, summarizing the results of this research and comparing them to the initial working hypothesis presented in the introduction.

Part I: Foundations & the Problem

This first part lays the conceptual groundwork for the rest of the thesis. The opening chapter addresses the metaphysical background needed to discuss personal identity over time, focusing on the debate between endurantism and perdurantism. Since the persistence of persons is inherently a temporal question, the metaphysical model we adopt for persistence will inevitably shape our understanding of identity. I argue that a perdurantist framework better supports a neo-Lockean theory of personal identity by allowing the mind to be temporally extended without presupposing identity across time.

The second chapter turns to the main theoretical proposals in the personal identity debate, focusing on the distinction between psychological continuity views and animalist approaches. While Parfit's reductionism and Lockean-inspired theories offer strong explanatory tools, they also face the challenge of circularity – especially when memory is used as a criterion for identity. I show that this issue is not resolved by merely shifting to a biological criterion, as animalism still depends on an account of mental unity to define the self-aware subject. This will lead us to the central problem: if both body and mind are necessary in some form, how can we articulate a theory of personal identity that avoids circularity while preserving explanatory power?

1 Metaphysics of Persistence

There are many things to say about personal identity, from a variety of perspectives. I believe this is the main reason why it seems so mysterious, and at the same time so exciting – not only to philosophers but also to anyone simply interested in human nature. However, to understand personal identity in the most rigorous and profound way, it is inevitable to first understand individuation in persistence. If neo-Lockeans are to argue for a personal identity grounded in memory, it seems that before addressing the nature of memory, we first need an account that captures how objects, including persons, persist through time. Moreover, to be compatible with memory-based and potentially hybrid theories, this diachronic account of personal identity must accommodate phenomena such as temporary psychological discontinuities as well as the constant change of the physical body.

Many have indeed expressed doubt about whether the metaphysical question of identity is a genuine and meaningful question in philosophy. From Hume's (2008) viewpoint, since there is no subject to which bundles of experiences are attached, there is no so-called self that persists through time; therefore, the idea of personal identity does not make much sense. For Wittgenstein (1958), our ordinary language demonstrates that it makes no sense to ask any question regarding the identity of something without context. For example, asking "Is this the same pen?" sounds absurd, because sameness and identity are relations an object has to itself. Furthermore, Frege (1948) argued that identity is a relation that makes no sense when applied to objects but only when applied to names with different senses (Textor, 2011). In that case, identity would not be a metaphysical problem but merely a linguistic one. Arguing in favor of psychological connectedness, Parfit (1984) claimed that strict identity does not matter when it comes to survival, and that survival is all that matters to us. That is, as long as psychological connectedness is maintained, it does not matter if strict identity becomes fuzzy. In the same line of reasoning, Johnson (2010) emphasized the importance of survival and ethical significance over a misplaced focus on theories of strict identity. Others, like Strawson

(1999), attribute our intuitive presupposition of the self to the phenomenology of personal experiences.

It seems that a clear answer to the nature of identity is impossible unless we apply a context. Regarding persons, philosophers have debated whether the mind or the body is better qualified to provide the right context for identity to take place. According to the skeptics mentioned above, the first step to answering the question about identity may be to let go of identity itself. If we reason with Parfit's survival view of personal identity, according to which survival is not understood in terms of strict identity, then there is no urgency in defining persons through a rigid framework of identity over time. Clinging to such a framework might be one reason why discussions often fall into circularities, as is the case with memory-based views of personal identity.

In this chapter, I take the question of identity as valid and examine the two dominant metaphysical views of persistence, which should illuminate our understanding of personal identity over time. These views concern how individuated objects persist through time: they either perdure or endure. This leads us to the perdurantism/endurantism debate. Persons can theoretically survive time either as four-dimensional creatures composed of temporal parts or as whole and complete creatures that exist wholly at each moment they traverse. In the end, I conclude that the question of identity over time is philosophically legitimate, since a meaningful distinction can be made between perduring and enduring objects. Although no conclusive answer is reached, perduring objects seem better aligned with what we expect persons to be as persisting creatures.

1.1 Persisting as Perduring or Enduring?

A direct way to distinguish between the two kinds of persistence is by considering the dimensions each takes into account. Endurantism, also known as three-dimensionalism, considers only the three spatial dimensions – width, height, and depth – hence its name. In contrast, perdurantism, or four-dimensionalism, adds a temporal dimension to these three spatial dimensions.

Endurantism always, in all circumstances, considers objects as their entirety. For example, a cat is, at every moment, a whole cat. This view faces certain difficulties, especially in accounting for the indiscernibility of parthood. For instance, it struggles to distinguish between the cat, the cat's tail, and the cat's ears. Suppose the cat was born complete and

whole but loses its tail to a horrible accident in year one (Y1). It seems that endurantism would be forced to claim that the complete cat before Y1 and the tailless cat after Y1 are not the same cat, which sounds absurd. How endurantism is unable to guarantee the discernibility of parthood is known as the 'temporal grounding problem' (Hawley, 2008), since the enduring object's identity at any given moment relies only on facts available at that specific moment. Perdurantism does not face this problem because it treats the temporary coincidence (before Y1) of the whole cat, the cat's tail, and the tailless cat as merely sharing temporal parts. Only the assembly of these parts forms the identity of the cat over time.

The sharing of temporal parts can be easily understood by analogy with the sharing of spatial parts: the upper half of a mountain and the lower half of the mountain share some of the same squares of soil, while other squares belong exclusively to either the upper or lower half. However, the upper and lower halves of the mountain are indiscernible not merely because they share some soil, but because they are physically part of the same mountain. Similarly, we do not question the indiscernibility of the cat and the cat's tail before Y1. However, as we will see shortly, perdurantism is not without its difficulties. The main challenge for perdurantism, as posed by endurantist proponents, is not merely to provide a positive account of persistence, but to demonstrate that endurantism fails to offer equally compelling explanations. In other words, the task is to show that perdurantism holds explanatory advantages over endurantism when it comes to accounting for persistence.

It is possible to visualize the difference between the two views of persistence by borrowing the spatial analogy to represent the dimension of time. If we cancel out the spatial dimension of 'depth' in both kinds of persistence, then three-dimensionalist objects can be fully represented in a flat image, whereas four-dimensionalist objects can only be fully represented through a 3D image composed of a sequence of images. This is similar to old cartoon motion pictures, where one drawing follows another nearly identical one – each slightly different – to render characters lively and in motion until the end of the film. Consider the example of a point as a persisting object in Figure 1. On the left side, the point is a three-dimensional enduring object: still, and existing as a whole point in the present. On the right side, the point is a four-dimensional perduring object that has instantaneous temporal parts – expressed by time slices just like the flat image of the enduring point – grouped together to form a whole. We can see a wavy line that represents the perduring point's motion in space

through time. This is how we may visualize that four-dimensionalist “[p]ersistence through time is like extension through space” (Sider, 1997: 197).

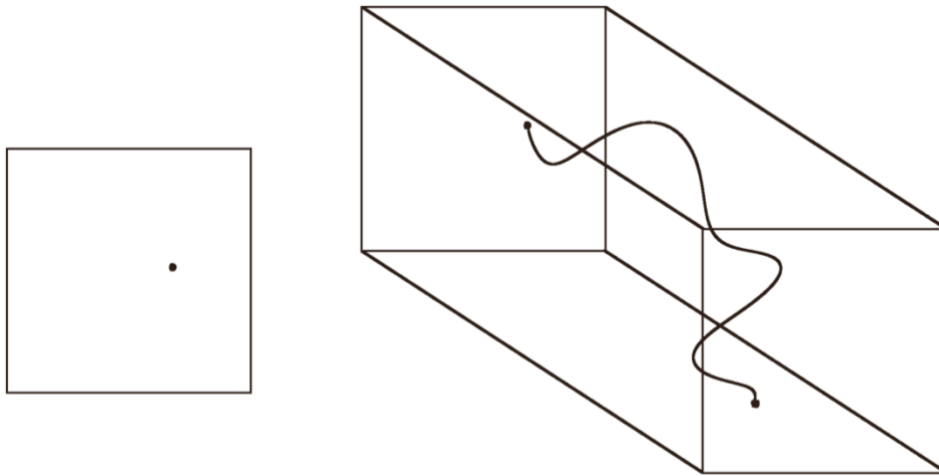


Figure 1 Three-dimensional point on the left; four-dimensional point on the right.

Representing the dimension of ‘time’ by borrowing the place of ‘depth’ in the diagram for the perduring point does not mean that the enduring point does not persist through time. It is simply a way of emphasizing the difference between how the two views understand persistence through time. Four-dimensionalist time slices, each containing a different version of the point, are the point’s temporal parts, which together form the larger ‘space-time worm’ that is the point. This is how perdurantism accounts for changes and temporal coincidences of mereological objects through the concept of four-dimensional worldtubes (Sider, 1997: 199–200).

Four-dimensional objects have a material composition, and although the visualization provides a straightforward distinction between the two views, it certainly does not capture the whole picture. For example, the illustration of the point cannot fully show the gaining and losing of past and future parts of the point as a persisting object, nor the parthood relations that constitute the mereological concept of perdurantism. Sider (1997: 200) provides the example of the tip of his fingernail. Even though today this tip is part of him, tomorrow, when he clips it off, it will no longer be part of him simply because that specific tip will no longer be a spatial part of him in future time slices. The four-dimensionalist considers the clipped-off tip not only compatible with the general mereological notion of “a part at a time” but also as an ‘atemporally exemplified property’ of Sider. This is because, even though the tip is no longer spatially part of him from the moment it is clipped off, as a past spatial part of his existing in past time slices, it remains part of the whole spatiotemporal worm that is Sider. It is not

because this property was only spatially distributed to Sider in the past that it is no longer part of him at a future instant. In this sense, it is an *atemporal part* of him. This idea of atemporal parthood is unacceptable to the three-dimensionalist, who considers properties such as “having that tip of the fingernail” and “being five feet tall” as temporal characteristics of an object. However, endurantism does not seem to offer a satisfying account of how these temporal characteristics cope with change through time. Atemporal properties are compatible with perdurantism because it regards each instantaneous temporal part of the whole as constituting properties of the whole: each time slice contains some spatial part that could be part of the entire space-time worm. The opposite is true for endurantism, which considers objects as wholly present at each time slice. Enduring objects are, in fact, also space-time worms (even though the term is more typically applied to perduring objects) that have no temporal parts but rather temporal extent, and hence the temporality of their characteristics and properties.

Some argue that the only logical difference between perdurance and endurance is the acceptance or denial of the inclusion of temporal parts in persisting objects. Because, just as perduring objects, enduring objects can also have *temporal extent* (Parsons, 2000; Hawley, 2008), the three- and four-dimensionalist debate is, in fact, quite meaningless (Hirsch, 1982; Magidor, 2016; Van Inwagen, 1981). However, the arguments from vagueness of composition and analyses of temporal grounding problems have shown that it is more likely than not that objects should have spatiotemporal parthood rather than be wholly present at each time. The vagueness of composition and temporal grounding problems highlight the indeterminacy of the indiscernibility of temporarily coincident objects. That is, before the cat loses its tail in Y1, the two views of persistence cannot agree on the relation between the cat and the tail when they temporarily coincide. Since perdurantism considers temporary coincidence as a “mere sharing of (temporal) parts” (Hawley, 2008: 197), it can respond to this problem quite straightforwardly by acknowledging the cat’s tail as one of the atemporal properties of the cat: one that the object, the cat, possesses through parthood from a certain time slice (its birth) to another (the precise moment of the cutoff during the tragic accident in Y1).

This advantage, nonetheless, does not seem conclusive, for endurantism can similarly claim that the temporary coincidence of enduring objects – just as that of perduring objects – survives change not through temporal parts but through spatial parts (Wasserman, 2002; Hawley, 2008). But if the analogy between temporal and spatial parts truly holds, why do more

philosophers seem to favor the perdurantist perspective of persistence, making it the ‘orthodox position’ (Oderberg, 1993: 63)? In the next section, I will argue, in line with Hawley (2008), that although perdurantism accounts for temporal coincidence similarly to endurantism, the four-dimensionalist framework ultimately offers more and is more suitable for the purpose of the present work.

1.2 Temporal Grounding of the Region

The central difference between endurantism and perdurantism, as mentioned above, is that while perduring objects exist only partially at a time, enduring objects are wholly present at each time. This distinction underlies a notion of ‘localness’ that must be considered: if an object is wholly present at a specific spatiotemporal region, at a particular point on a time slice, then fundamental facts about that object should also be confined to that region. Since perduring objects do not exist wholly in any single region but only partially, they can have fundamental facts ‘distributed’ across different regions where the object is partially present.

To better define the subtle difference between enduring and perduring objects, Hawley (2008) further distinguishes two kinds of temporal grounding problems: one is *object-focused*, concerning the temporary indiscernibility of the object’s basic properties; the other is *region-focused*, concerning the temporary indiscernibility of regional properties, such as temporary spatial coincidence and mereological identity.

Let’s first consider the region-focused temporal grounding problem. Perdurantism is especially good at responding to classical puzzle cases of personal identity, including fission, fusion, and reduplication, by adopting the ‘multiple occupancy’ thesis (Noonan, 2019: 13–14; Hawley, 2008: 203–204). Consider a fission case where P_2 , P_3 , P_4 , and P_5 are four distinct post-fission people existing at time t_2 , and they are all products of a fission that happened to their common ancestor person, P_1 , at time t_1 . To account for the existence of P_2 , P_3 , P_4 , and P_5 , perdurantism can adopt the view that these are not brand-new people who have come into existence since t_2 , but rather people who have always existed since the beginning of P_1 ’s existence: they have all been sharing continuous spatiotemporal regions, i.e., *temporally coincident*, all the way up to t_2 . This view does not violate the ‘only-x-and-y’ principle and can account for puzzle cases. This is because how it considers two people to be distinct at a certain time does not rely *intrinsically* on the region but rather on facts that are *extrinsic* to it (Robinson, 1958; Lewis, 1983). To borrow Lewis’s (1983) expression, when P_2 , P_3 , P_4 , and P_5

are in temporal coincidence, they are four minds but in a single thought. The multiple occupancy view can not only account for logically possible situations in puzzle cases but also clinical ones, such as individuals with dissociative identity disorder (DID).

DID is a characteristic phenomenon among dissociative disorders. Individuals who experience DID undergo a “disruption of identity characterized by two or more distinct personality states”, and this disruption “in identity involves marked discontinuity in sense of self and sense of agency” (APA, 2013: 292; Kind, 2020: 33). Again, in Lewis’s expression, this situation is analogous to multiple occupancy, except that this time, *P2*, *P3*, *P4*, and *P5* fail to share a single thought and instead have their own apparent thoughts and streams of consciousness. DID is a real-life fission of consciousness confined within an individual’s singular body. The DID case seems to not only provide advantages to the perdurantist view but also shed light on the relation between psychological continuity and bodily conditions regarding personal identity. However, that is a discussion for future chapters. For now, the case should be useful simply to point out that the perdurantist ontology of persons can account not only for puzzle cases but also for similar cases in real life.

Wasserman (2002) points out that it is fairly straightforward to illustrate how perdurantism can account for multiple occupancy; the difficulty lies rather in explaining how endurantism fails at this task. As we have seen, perduring objects can be determined not only spatially but also by temporally extrinsic facts. Similarly, enduring objects can indeed be determined by spatially extrinsic, but only temporally intrinsic, facts located at a specific time slice, since they exist wholly at each instant. This creates a difficulty for endurantists regarding determination: it is incompatible with the claim that temporally extrinsic facts, events occurring in distant spatiotemporal regions, should determine whether an object is wholly present at a spatial instant (Hawley, 2008: 205). In other words, temporal extension does not seem as compatible with enduring objects as endurantists claim, since the question of “how many objects are in this specific spatiotemporal region” is in fact a temporally intrinsic matter. This is what makes temporal grounding a serious problem for endurantism, as it cannot adequately account for certain puzzle cases. When *P2*, *P3*, *P4*, and *P5* come into existence at *t2*, endurantists are obliged to consider them as newborns and announce the death of *P1*, even if *P2*, *P3*, *P4*, and *P5* might have psychological continuity with *P1* (memory-based theories) or share the exact same bodily components (body-based theories).

The endurantist conclusion not only sounds counterintuitive but also highlights a clear distinction between the two ontological views of persistence, providing strong evidence of the perdurantist advantage regarding the temporal grounding problem of regions. Perdurantism straightforwardly accounts for minor changes within each spatiotemporal region throughout the entirety of an object's persisting lifetime. This view also aligns with our intuitive recognition of *P1* as a single object, even while simultaneously acknowledging the existence of *P2*, *P3*, *P4*, and *P5*. As Lewis suggests, in such cases, we are effectively counting in terms of common parthood rather than strict identity (Lewis, 1983; Hawley, 2008).

1.3 Temporal Grounding of the Object

The object-focused temporal grounding problem concerns the temporary indiscernibility of an object's 'basic properties' (Hawley, 2008). This problem was notably proposed by McGrath (2007) as a challenge to perdurantism, where an object's 'basic properties' at a given time are contrasted with both its historical and future properties. The claim is that if two or more objects share spatial parts at a time slice, they also share all of their basic properties at that instant. This argument opposes the multiple occupancy view because it interprets spatiotemporal parthood solely in terms of the basic properties of a single object, thereby undermining perdurantist mereology. Moreover, through the principle of 'causal determinism', it rejects the possibility of temporary coincidence between two or more distinct objects (Hawley, 2008: 207): if any group of things coincides temporarily, it is only their basic properties that constitute the object. This causal determinism presupposes a natural law whereby past and future properties are determined by current basic properties.

According to McGrath's analysis, both endurantism and perdurantism face similar challenges in accounting for the temporary coincidence of multiple objects. Among the various factors involved in an object's basic and futural properties, the 'sortal property' – which defines the kind of object – stands out as particularly fundamental. For example, a cat and its tail share basic properties when they coincide at a time slice, but they differ specifically in kind. That is, the cat and its tail respond differently to changes because they belong to distinct sortal categories. This argument complicates the idea that perdurantism holds any clear advantage over endurantism. However, Hawley (2008) points out that while both views may be equally capable of accounting for temporal coincidence among objects of the same kind, perdurantism handles temporal coincidence of objects of different kinds with greater

ease. The explanatory power of each view in these two contexts will be assessed by their reducibility in what follows.

It is, in fact, not complicated to see how, according to both endurantism and perdurantism, *same-kind* temporal coincidence is irreducible. For example, consider a lump of clay that is cut in half, producing two separate lumps. Neither endurantism nor perdurantism can fully explain how this original lump becomes these two distinct lumps. Although perdurantist proponents might intuitively want to argue otherwise, during the temporal coincidence of the pre-cut half-lump, there is no way to know which clay particles will belong to this later half-lump or that later half-lump.

However, the situation differs when it comes to temporal coincidence of *different-kind* objects: perdurantism appears better able to reduce facts about such circumstances than endurantism. For example, consider a statue that is destroyed and replaced by a large, squashed lump of clay. Within the mass of the squashed lump, we can identify a lump that was the very same lump present when the statue was intact. After the squash, only the lump survives as its kind (the lump of clay) while the statue no longer exists as its kind. Nonetheless, before the squash, there was indeed temporal coincidence between the lump of clay and the statue. The challenge then becomes determining the identity of the two during this temporal coincidence: “Is this object a statue because of its future career, or does it have the future career it does because it is a statue?” (Hawley, 2008: 209).

Endurantism can only accept the second claim, whereas perdurantism tends to adopt the first. Since enduring objects exist wholly at each instant, endurantism cannot mobilize futural properties to help determine their basic properties in the current spatiotemporal region. It is therefore obliged to take the coinciding lump of clay merely as a basic property of the statue at that moment – it is a temporary property specific to that time slice (or block of multiple time slices). The sort, or kind, of the enduring object can only be determined by the basic properties available as local data within the spatiotemporal region. Conversely, perdurantist objects have temporally distributed futural properties at their disposal: the sort, or kind, of the object at a time slice can be determined by properties located outside the current time slice. In this example, where endurantism cannot distinguish the lump of clay from the statue, perdurantism considers two objects of different kinds to coexist in one temporal coincidence.

We have seen that the metaphysics of persistence is complex, and both endurantist and perdurantist views capture fundamental aspects of its nature. However, when considering the temporal grounding problem – both from regional and object-focused perspectives – four-dimensional perdurantism appears better suited to accommodate our intuitions about the determination of object identity over time.

1.4 Conclusion

In this chapter, I have examined the two main metaphysical views of persistence. Despite lingering doubts about the genuineness and significance of the identity problem, I have argued for a clear distinction between endurantism and perdurantism, ultimately favoring the latter to underscore the importance of addressing the problem of identity over time.

The analysis of temporal grounding problems suggests that while endurantism accounts for certain cases, perdurantism applies to a wider range. That is, one might consider that all perduring objects could be seen as enduring objects: if we disregard the succession of time slices. Endurantism would hold true if the world were atemporal, if each object were determined solely by its contemporary basic properties, and if nothing more were needed. The flaw in this view lies in its requirement for the presence of the whole, which ties the concept of objects to a rigid framework of identity: one based on absolute sameness. We may therefore conclude that perdurantism is more advantageous in explaining persistence due to its ability to accommodate constant change: it captures identity through time by appealing to similarity, rather than strict identity, between an object's spatiotemporal parts. We might even boldly adopt the idea of identity-less-ness between an object's temporal parts in a diachronic context as the only coherent way to account for identity over time.

If perdurantism is true, personal identity becomes a temporally contextual concept, which opens up the possibility that there may be necessary but not sufficient criteria for personal identity. Even though, under normal circumstances, we all know – one way or another – about our own pasts, there is no way of knowing what will happen to us in the future. Therefore, there are temporally extrinsic facts unavailable to me for determining who I am at the current time slice. For example, I may never know whether I am, in fact, two different persons at present because I may never know, at present, whether I will eventually undergo reduplication (a form of fission), if teleportation technology succeeds in creating a

post-teleportation version of me at the destination without erasing the pre-teleportation version (Parfit, 1984).

However, the fact that four-dimensionalism is a more materialistic concept does not imply that it favors bodily or physical persistence. Although consciousness and memory are immaterial concepts and may not, *prima facie*, seem compatible with perdurantism, this contrast is misleading. A stream of consciousness captured at any one time should represent the entirety of that consciousness. It seems absurd to claim that a stream of consciousness is composed of temporal parts or “time slices”, as though these were mere fragments of the stream as a whole.

As Shoemaker (2012: 133) points out, persons can indeed have stages (person-stages) without committing to the view that these stages are mere temporal parts. A complete person exists at each of these stages. The idea that persons endure rather than perdure seems to align more closely with Locke’s claim that “person” is a forensic term, raising the concern that a perduring person may not be fully responsible for their actions. However, this line of reasoning is ultimately misleading. Perduring objects can account for change between stages while preserving identity, and this is not incompatible with responsibility or agency. Still, it is admittedly less straightforward to see how memory-based persons might perdure.

Both perdurantism and endurantism can account for bodily and psychological persistence, as they are views about how objects persist through time, rather than about the nature of physical bodies or mental states. It is not necessarily the case that perdurantism better explains bodily persistence and endurantism better explains psychological persistence, or vice versa. Unlike persons, human bodies, in Locke’s sense, can more easily be understood as having both temporal and spatial parts. Our bodies not only occupy space at each time slice but also possess different properties at different moments. In fact, consciousness itself can have properties and be associated with varying characteristics across different time slices.

Perdurantists do not take person-stages to be persons in themselves, but view the temporally extended series of stages as constituting a person. The entirety of these stages constitutes the numerical identity of the person. By contrast, endurantists regard each person-stage as the same whole person persisting through time. The endurantist view may seem more straightforwardly compatible with certain moral intuitions, as it posits an unbroken identity across time. However, the perdurantist framework offers a more dynamic ontology of persons, one that reflects how mental life unfolds. While this does not, by itself,

directly establish a better account of personal identity, it suggests that perdurantism is better suited to capture the temporal structure of psychological phenomena. For instance, mental properties such as intentions, experiences, and self-representations are not fixed at a single moment but develop and interrelate across time. The simulation theory of memory strengthens this point by showing that remembering involves an active reconstruction of past experiences from the present, projecting the self across time without requiring a continuous causal chain between stages. This fits naturally with the perdurantist model, where psychological continuity can be understood in terms of temporally distributed mental stages. A hybrid version of neo-Lockean theory that prioritizes memory over bodily continuity may thus find its most coherent expression within a perdurantist framework.

Before turning to memory theories in the next chapter, I will examine several prominent theories of personal identity and evaluate their compatibility with what appears to be the more promising view of persistence: perdurantism.

2 Personal Identity Theories

What am I? What are we? What is the thing that is asking all these questions? Is there such a thing? The question of personal identity is ancient, and the arguments for different views are abundant. The term “personal identity” has been so widely used that it has become a philosophical jargon referring to quite a lot of different things (Madden, 2024). In the previous chapter, we discussed the metaphysics of persistence and how personal identity over time seems to be structured. To assess whether memory can constitute personal identity without circularity, we must consider rival accounts and their compatibility not only with persistence but also with the memory account. In the present chapter, I will provide an overview of theories on the constitution: the metaphysics of the person. That is, the question of what constitutes persons: what makes a person.

For example, this research has been motivated by the neo-Lockean contention that memory, and more generally ‘consciousness’ and ‘conscience’, are what constitute persons. However, despite the primary focus on memory-based theories of persons, it is important to also consider body-based and hybrid theories for a more comprehensive comparison. To build a sound and compelling theory of personal identity, then, is to construct one that simultaneously incorporates an ontological account of metaphysical persons and a theory of persistence to explain numerical identity over time.

What distinguishes persons from all other organisms on Earth appears to be their uniquely developed capacity for mental life (Locke, 2008; Noonan, 2019; Kind, 2020; Madden, 2024). People manifest themselves in the world through their bodies – through decisions, actions, and expressions – which in turn are shaped by what they have learned, what they believe, and what they are capable of imagining. In this sense, persons are deeply constituted by their minds. However, being a person is not reducible to how one manifests oneself; a person can be a mind without necessarily being the self that is expressed at a given moment. This suggests that the discussion over the constitution of persons runs parallel to that of the constitution of minds: a point already present in Locke’s (2008) use of both ‘person’ and ‘self’ to describe “a being with a reflective perspective on its own mind” (Madden, 2024: 1). This

perspective need not imply a dualist bias. Rather, it should be seen as neutral: the mind's constitution might be entirely independent of the body, as dualists claim, or entirely or partially dependent on it, as physicalism and hybrid theories respectively maintain.

In line with the contemporary empirical spirit and the aim of gathering evidence about the constitution of persons, it might now seem reasonably straightforward to turn directly to what we perceive as a person in everyday life: the person's entire body: the perceivable identity to others. However, Cartesian dualism and certain earlier religious perspectives have led many philosophers, through reflections on the relationship between body and mind, to consider the psychological aspect as more fundamental than the physical. In recent years, this trend has shifted with the rise of influential theories such as animalism (Noonan, 2019; Lucci, 2021; see Olson, 2003; Bailey, 2015). Philosophers of personal identity have begun to favor either fundamentally body-based accounts or hybrid theories that aim to reconcile body and mind. A central debate within these approaches concerns which aspect – body or mind – should be given a priori foundational status. Among these theories of personhood, some appear more readily compatible with perdurantist accounts of persistence than others, though each faces its own set of challenges. In the end, I conclude that a version of hybrid theory offers the most promising solution.

2.1 Memory-Based Theories

The discussion about what makes a person a person relies heavily on observable evidence, from which we can construct a set of criteria and then determine which of these are necessary or sufficient for something to qualify as a person. The most intuitive evidence we might find for the existence of a person is the evidence of our own existence, since that is what we take ourselves to be: persons. This is how Descartes arrived at his famous statement: *Cogito, ergo sum* – “I think, therefore I am”. It seems plainly obvious that if I am able to entertain the thoughts currently present in my psychological environment, there must be something that is doing the thinking – an agent – and what could that subject be, if not a person who happens to be me?

Hume (2008), however, rejects the idea that there is a metaphysical entity such as the self or subject that serves as the agent entertaining thoughts. Instead, he argues that the thoughts we observe through introspection are merely causal reflections of the world, making a person a bundle or collection of different perceptions and experiences. This Humean idea

2 Personal Identity Theories

naturally inspires an identity-less way of accounting for the sameness of a person over time: that there is, in fact, no metaphysical self, and that personal identity is instead constituted by mental experiences in interaction with the physical world.

However, the textbook reference – the most influential work in the field of personal identity that has endured to this day – is Locke's, which inherits a more nuanced version of Cartesian dualism. Locke does consider a person to be a metaphysical subject capable of agency through the mind, but he also describes an infant's mind as a piece of white paper that accumulates content over time, which seems to echo the idea of a bundle of experiences that Hume refers to.

According to Locke, the contents accumulating on the "blank paper" analogy correspond to how persons are constituted by their memories or stream of consciousness. That is, regardless of any physical element, if the continuity of memory is interrupted, then the same person no longer exists. This is a strong claim, which we can test within the four-dimensionalist perdurantist framework. If the forward-moving point in the time dimension represents the stream of consciousness, then each time the person falls asleep or forgets something, that point disappears from the spatiotemporal prism and reappears when the person wakes up or repairs the lost memory. This creates a problem of discontinuity that the original Lockean view struggles to account for.

The original Lockean view seems reasonable in some respects; for example, it doesn't seem natural or fair to say a person remains the same after severe trauma causing total amnesia, which could lead to dramatic character changes and other psychological effects. Nonetheless, the claim is too strict, as it cannot account for very common phenomena such as forgetting early childhood memories or going to sleep every night. For this reason, Locke's sympathizers have elaborated the memory criterion into either 'psychological connectedness' or 'psychological continuity'. This means that as long as there is continuity in certain mental states – such as beliefs, desires, and memories – the person exists and continues to persist over time. When there is overlap in some psychological features between two person-stages, a psychological chain is formed, creating psychological continuity that constitutes one and the same person. If Locke's memory criterion can be summarized as "*P2* at *t2* is the same person as *P1* at *t1* if and only if *P2* at *t2* is linked by continuity of [memory] to *P1* at *t1*", then psychological continuity can be formalized as "*P2* at *t2* is the same person as *P1* at *t1* if and only if *P2* at *t2* is psychologically continuous with *P1* at *t1*" (Noonan, 2019: 10–11).

The most prominent objection to Lockean and other neo-Lockean accounts is the circularity objection, first raised by Butler and Reid: memory presupposes personal identity and therefore cannot constitute it. Whether memory can be identity-less, and thus can constitute personal identity in a non-circular way, is the subject of later chapters. For now, we will consider this objection in light of the traditional definition of memory – that we remember correctly and genuinely only because I am the same person as the one who experienced the event I am recalling.

In fact, the circularity problem applies not only to memory but to all mental phenomena. Whether it's memory, perception, action, sensation, emotion, desire, thought, or imagination, it is traditionally believed that these are properties belonging to an agent or subject, often characterized as the self. This common belief in the 'ontological dependence' of mental states on a self – the idea that such experiences can only occur if there is a subject who has them, that is, a particular 'bearer' of these experiences – is closely tied to the way ordinary language expresses these experiences (Madden, 2024). For example, "there is a cat that grins" makes sense, as do "there's the cat's grin" and "there's the grinning cat". Nonetheless, saying "there's the grin" on its own doesn't seem to mean much; it prompts us to ask, "Whose grin?". Since, grammatically, it is difficult to refer to experiences as properties without mentioning a subject, we intuitively consider that experiences are ontologically dependent on such a subject – the bearer of those experiences. If mental experiences can only occur when there is a present subject, then the subject (the person) constitutes those mental experiences.¹⁰ Therefore, mental phenomena cannot, in turn, constitute personal identity. This is how the circularity objection also applies to the broader psychological continuity theory of persons.

The other objection to memory-based theories is methodological. Descartes, Hume, and Locke's views were all "arrived at 'from the armchair'" (Madden, 2024: 4), meaning they are purely theoretical, based on intuitions and suppositions from introspection – 'from within' – and lack empirical support as a foundation. Therefore, these views cannot truly reflect the reality of what persons are. Taking my own thoughts as evidence seems obvious only to myself

¹⁰ However, there is a line of argument that remains skeptical about the connection between experiences and bearers. For example, Anscombe explores, through a linguistic lens, the idea that there is no self as the object of first-person usage in ordinary language. This argument will be further developed in Chapter 4, along with her view on action.

2 Personal Identity Theories

since these thoughts are internal and, although intrinsic, provide no observable evidence of my person to a third party. This is the virtue of empirical studies: their objects are not only observable but also testable and analyzable. This presents the main difficulty for mind-based theories, since mental states are invisible and, at best, only phenomenal from the inside.

As we will see next, both the challenges of circularity and methodology faced by memory-based theories can be readily addressed by body-based views of personal identity. However, these views have their own problems as well.

2.2 Body-Based Theories

Recent influential body-based views stem from animalism, a position that answers the question “What am I?” with the reply that we are biological organisms: “human persons are one and the same as biological organisms of a certain kind” (Madden, 2017: 2; see also van Inwagen, 1990; Olson, 1997; Snowdon, 2014). On this view, persons *are* their bodies. Intuitively, this may seem easier to conceptualize than, for instance, the soul, whose metaphysical status is unverifiable and empirically unsupported. Moreover, some have seen body-based theories as a way to sidestep the circularity objection facing memory-based accounts. If the circularity arises from assuming a subject that bears mental experiences, then one might be tempted to deny that such a bearer exists, suggesting instead that persons are constituted by mental experiences themselves. But this leads to serious difficulties, as it is unclear how experiences could exist without a unifying subject. In response to this problem, body-based theories offer an alternative: grounding personal identity not in mental states or experiences, but in the continuity of a living biological organism.

Before proceeding to critique body-based theories, it should be acknowledged that my body is the most obvious evidence of my existence. Expressed in terms of persistence, this becomes the claim that “*P*2 at time *t*2 is the same person as *P*1 at time *t*1 if and only if *P*2 has the same body as *P*1 had” (Noonan, 2019: 2). However, defining persistence in this way aligns persons with physical objects like bridges or oak trees – entities that persist physically, not mentally or otherwise. While such a view may seem counterintuitive within the dualist tradition, it has gained significant traction in recent decades. Olson (1997, 2002, 2003) articulates this position as ‘animalism’, arguing that persons are animals. This biological approach aligns well with everyday experience: in normal situations, we rarely question whether someone is the same person if they have the same body. However, many

philosophers find it hard to accept that human persons are merely human animals – It appears to overlook something essential, namely, our intellectual and mental life.

A deeper issue arises when considering how the body might account not only for physical continuity but also for mental experience. While the body functions as a ‘causal interface’ between ourselves and the world, mental experiences seem to require a metaphysical bearer. Causal ownership of properties can be transferred, but metaphysical ownership cannot: “if it is part of the nature of an experience to be related to a particular thing, then it couldn’t possibly fail to be related to that thing” (Madden, 2024: 9). This suggests two possibilities: either one adopts a Humean no-bearer view, where mental phenomena need not belong to anything; or body-based theories must explain how mental phenomena are physically located or locked into the body.

Only a strong form of animalism insists that persons are necessarily and sufficiently constituted by their entire body. But this claim faces serious objections. For instance, it seems unjust to say that a man who loses his legs is no longer the same person he was before the accident. This point generalizes: I could plant an oak sprout before moving out of my childhood home, and years later, recognize the grown oak tree in the yard. Despite the changes in appearance, it would seem strange to deny it is the same tree. Moreover, the human body undergoes constant change, so identifying persons with a fully preserved body is implausible. Thus, the body may be characteristic, but not sufficiently constitutive of personal identity.

The relevance of this point has been explored by Shoemaker (2003a) and Parfit (1984) through the famous Brownson case. In this case, Brown and Robinson are two individuals. Brown’s brain – carrying all his memories, thoughts, and character – is transplanted into Robinson’s body, while the rest is discarded. Most would intuitively judge the resulting person to be Brown, not Robinson. For memory theorists, this thought experiment disqualifies a bodily criterion and privileges psychological continuity. To respond, the strong animalist might retreat to a weaker position, reducing the amount of body required to constitute a person – perhaps down to just the brain. This strategy, grounded in empirical neuropsychology, appears attractive: it retains verifiable biological continuity while preserving the organ responsible for psychological life. However, if the brain becomes the new bodily criterion, the notion becomes less radical. For example, if personal identity resides in the brain, wouldn’t it be possible to reduce it even further – to parts of the brain or just certain neural systems in the brain? Indeed, empirical studies (Sperry, 1968; Nagel, 1971; Halder et al., 2021) show that

2 Personal Identity Theories

the presence of both hemispheres isn't necessary for consciousness, as each hemisphere can be conscious independently. This suggests that even the whole brain is not strictly required for personhood, further weakening the bodily criterion.

Even this minimal version of the bodily criterion is challenged by Williams (1973b). The Brownson case and other similar thought experiments rely on preserving the physical part that houses psychological identity. But Williams introduces an imaginary device capable of storing all the information from a person's brain. If we extract Brown's brain after transferring all its information into this external device, and then transplant the brain into Robinson's body, the resulting being can no longer clearly be identified as Brown. The brain has become no more than meat, with its content already offloaded elsewhere. This undermines the idea that psychological identity is tied to any particular biological part. If it is logically conceivable for all mental content to be stored externally, then the body – whether whole or in part – cannot be a sufficient basis for personal identity. While Williams does not commit to a radically body-based view, he leans toward the psychological continuity account, offering compelling arguments against the sufficiency of any pure body theory.

We have seen how, by nature, body-based views can overcome the two main challenges faced by mind-based views. There is no circularity in defining personal identity by the body, since you don't need to be a particular person to have the body you have. The constitution is one-way only. Regarding the problem of lacking an observable and analyzable object as evidence, the body (or parts of it) can be straightforwardly examined empirically. However, body-based theories are not without difficulties. Their main issue is that they cannot preserve the primary virtue of mind-based views: accounting for mental activities that are unique to human beings – characteristics that set us apart from other living things, including animals. In this case, considering a framework that preserves the virtues of both views may provide us with a winning solution.

2.3 Hybrid Theories

At the current stage of theoretical development, few would be stubborn or reluctant enough to commit exclusively to either extreme view – that persons are made solely of mind or solely of body.

On a spectrum ranging from strong body-based to strong memory-based views, we can identify 'strong animalism' at one end and the 'strong psychological view' at the other.

Identity-Less Personal Past

Theories that claim the body is the necessary and sufficient criterion for personal identity fall into the first category, while views such as Locke’s original proposal – which holds that memory or consciousness alone is both necessary and sufficient for personal identity – belong to the second.

However, in contemporary literature, instead of treating them apart, debates tend to focus more on the relative importance of both mental and body criteria and the proper relationship between them. Participants in these debates are often considered hybrid theorists of personal identity. Among them, those who advocate for a complete and more stringent criterion based on the human body are known as ‘weak animalists’. In contrast, descendants of mind-based theories who accept certain conditions – such as a causal connection to physical parts of a person (Shoemaker, 2003a) or the idea that only the brain suffices for personhood as manifested by the mind (Parfit, 2012) – are classified as proponents of the ‘weak psychological view’.

Table 1 Personal identity accounts on a body-psychology spectrum.

<i>Views</i>	<i>Core Idea</i>	<i>Example Case</i>	<i>Example Proponent</i>
Strong Animalism	Personal identity equals biological continuity.	I can survive complete amnesia; I am the same person even if I lose all my psychological traits, as long as I still have the same body.	Olson, Snowdon, Van Inwagen.
Hybrid Theories: Weak Animalism / Weak Psychological View	Personal identity is grounded in causal biological continuity, but psychological factors matter pragmatically. It can also be preserved through branching, duplication, or quasi-memory.	If the rest isn’t discarded in the Brownson case, the scenario becomes one of brain swapping. If the brain’s content is stored in a special device, then the resulting products are merely bodies: Brown’s body with Robinson’s brain is Brown, and Robinson’s body with Brown’s brain is Robinson. If I undergo fission, fusion, or reduplication, numerical identity becomes fuzzy; but what matters is that I survive through my psychologically connected successors. In our empirical world, the brain enables this continuity.	Madden, Parfit, Shoemaker, Williams.
Strong Psychological View	Personal identity requires an unbroken chain of mental states.	If my psychology is transferred into Brown’s body, then what physically looks like Brown is, in fact, me. My personal identity follows the transfer of my mind into Brown’s body.	Locke

2 Personal Identity Theories

The most interesting debate lies along the hybrid theories line (in Table 1), which includes both ‘weak animalism’ and the ‘weak psychological view’. How much of the body should be accepted as a necessary condition? How strict must psychological continuity be to count as a necessary condition? What would a combination of the two criteria look like – one that could be considered sufficient to fully determine personal identity and best answer the questions “What am I?” and “What are we?”.

Just as the debate between perdurantism and endurantism centers on the temporal mereology of persisting objects, the debate between body- and mind-based theories can be seen as involving the spatial mereology of persons. Madden (2017) argues against Parfit’s (2012) position, claiming that persons cannot be merely the brain but must instead be identified with the entire body – the complete human animal. Parfit’s mereological view is rooted in his survival-based account of personal identity. According to Parfit (1984), what matters in survival is not strict numerical identity, but rather the continued existence of one’s mental properties in another being.¹¹ Even if identity becomes indeterminate in cases like fission or fusion, survival is preserved if someone carries on my psychological traits. Though this is a simplification of Parfit’s nuanced view, it captures his general stance in response to objections. Because of this, he is willing to accept a bodily criterion, but reduces it to just the brain. While Parfit acknowledges that we are human animals in a biological sense – that our bodies make us similar to other animals – he denies that the whole animal is necessary for personal identity. As long as the essential part, namely the brain, is preserved, personal identity can continue. This forms the basis of his mereological ‘embodied parts’ view: the claim that, the persons we are, are in fact physically smaller than the biological individuals we are (Madden, 2017: 3; Parfit, 2012: 14).

The problem that Madden identifies in Parfit’s view is that if we accept that both the person and their parts are capable of thinking (despite the distinctions Parfit draws between derivative and non-derivative, direct and indirect, and conscious and unconscious thinking), we face the obvious objection of too many thinkers, which cannot be reconciled with our ordinary linguistic uses of the first person. First-person thoughts involving body parts are particularly problematic, such as “I have two arms and two legs”, or those involving bodily

¹¹ This sounds particularly similar to Hume’s view.

sensations, like “I feel pain in my left arm”. These ‘I’-claims require the referenced body parts to stand in the correct parthood relation to what is being referred to by ‘I’ – a determinately individuated entity. This implies that, if it can be shown that animals, rather than brains, are the genuine thinkers, then animalism is correct. Conversely, if brains alone can be shown to think, then the mereological brain-based view of personal identity is vindicated.

By differentiating the thinking done by the brain from that done by the other parts that make up the whole animal, Parfit elaborates the notions of derivativeness, directness, and strictness of thinking. The non-derivative thinker is the brain, which thinks directly in a “strict, first-rate sense”, whereas the derivative thinker is the rest of the animal’s body, which thinks only indirectly, in a “loose, second-rate sense” (Parfit, 2012: 21–26). Having drawn this distinction, Parfit responds to the first-person linguistic objection on two levels.

The first is a seemingly bold claim about the meaning of ‘I’: that first-person thoughts, in fact, “embody a restriction to refer only to non-derivative thinkers” – that is, the brain (Madden, 2017: 7; Parfit, 2012: 18–19). The second, which Madden considers worthy of deeper exploration, is a claim about reconstructing the referent of ‘I’. Parfit agrees that in ordinary language, ‘I’ does not strictly refer to the non-derivative thinker. However, he invites readers to use ‘I’ henceforth to refer explicitly to the non-derivative thinker, so that we might claim that, rather than being the entire animal, we are instead “the most important parts of these animals” (Parfit, 2012: 24–25).

Although Madden finds the general direction of Parfit’s reconstructive strategy rather attractive, he considers Parfit’s invitation odd. Animalists, after all, could just as easily offer a similar suggestion in their favor – encouraging readers to use ‘I’ to refer explicitly to the derivative animal thinker, rather than the non-derivative one – so that we may truly claim that we are the animals we are, rather than mere parts of those animals. This, Madden argues, renders Parfit’s distinction between derivative and non-derivative thinking ineffective in answering the question “What are we?”.

Instead, Madden (2017: 8) proposes a different distinction: between what is self-aware and what is not. The concern about what persons are, he claims, “is not simply a concern about which thing is thinking these thoughts, and having these experiences”, but about “which thing is thinking these thoughts, and having these experiences, *in a self-aware way*” (Madden, 2017: 8, emphasis added). He argues that the unique self-aware thinker is the animal, even if there are multiple thinkers within it.

2 Personal Identity Theories

Contrary to what our intuitions might suggest when applying the term ‘self-awareness’, Madden’s view is arguably compatible with both Humean skepticism and Anscombe’s objectless concept of the ‘I’ in ordinary language: the idea that when we introspect our mental phenomena, there is no self to be found, and that the use of ‘I’ fails to refer to, and justify the existence of, any object that is supposedly the self.

Animalism can advance the argument that there is one self-aware thinker in the animal, and that “the animal is uniquely self-presented”, enabling ‘reflexive channels’ that are non-introspective (Madden, 2017: 9–10). These reflexive, non-introspective channels do not provide information about mental properties; rather, they are monitoring systems for bodily sensations such as the sense of balance, visual self-location, and proprioception. According to functionalism, these sensory channels evolved to monitor the bodily states of the animal as a whole organism. These channels form a complex system through which the organism sends signals to the brain, which does the thinking. However, the organism itself is not the thinking brain, but the entire complex being that is numerically distinct from the brain. In this sense, it is the organism that is self-aware, not the brain.

Madden’s central argument is that persons are their human animals because they cannot be reduced to the brain. While the brain may have the special function of establishing relational properties between different body parts and their sensory apparatus, it is not the brain that is self-aware, but the animal as a whole. Without the bodily apparatus of the human animal, the brain would be unable to obtain any information through reflexive channels, and it would lose its function, becoming inert, like meat. In this sense, animalism should be considered true: the body of the person – understood as a complete organism, a living being – is fundamental, because self-awareness depends on it.

In Madden’s line of argument, it seems possible to go further and argue that even though it is not the brain itself that is self-aware, but rather the organism as a whole – possessing reflexive channels of awareness – the brain remains essential as an interpretive medium. Without the brain, even if these reflexive channels carry information (from a limb’s sensory apparatus, for example), there would be no destination for this information to be processed. The sensory input alone would remain a solitary property, uncontextualized and uninterpreted.

To illustrate this, imagine a tragic individual who, over the course of several years, suffers a series of severe accidents. In year *Y1*, they lose their left leg; in year *Y2*, their right

leg; and by year *Y4*, they no longer possess any limbs – yet they remain alive, with a perfectly intact brain. The sensory apparatuses of these limbs are gone, and the reflexive channels are no longer transmitting data to the brain. Does this mean the individual is numerically no longer the same person as they were in *Y1*, when all limbs were intact?

We can push this thought experiment even further. Suppose the individual's heart is replaced with a perfectly functioning artificial heart. As long as the reflexive connection between this artificial organ and the brain is preserved, the organism continues to live. If we accept that, then we could imagine replacing the entire body with a precise simulation of the original organism – something akin to the imaginary device Williams envisioned for storing brain information, but extended to the whole body – as long as it retains the capacity to maintain reflexive relations with the brain. Why not, then, go a step further and replace also the brain itself with a machine that perfectly replicates its functions (i.e., exactly like Williams's device)?

In such a case, both the biological body and the artificial simulation appear to function as interfaces with the world. What truly constitutes personal continuity, it seems, is the accumulation and processing of experiences – regardless of whether the brain is made of meat or machine. This leads us to a Humean and Anscombean conception of the person: selfless, dependent on mental phenomena without requiring a metaphysical “self”.

This imagined case weakens Madden's argument that personal identity requires the entirety of the human organism, as opposed to Parfit's mereological view centered on the brain. However, Madden remains right to stress the importance of the connection between our mental states and our physical apparatus. The case of the tragic individual does not prove that the person persists through the animal body alone, but it does show that self-awareness – indeed, personhood itself – depends on a functioning interface with the world.

Whether this interface is biological or artificial, persons require some material embodiment to be considered existent. This safeguards us from falling into the trap of Cartesian dualism, where immaterial thoughts or souls might exist without observable anchors. Ultimately, the question “What are we?” can only be answered by acknowledging that, even if we can conceptually imagine disembodied thoughts or souls, such entities do not fit our ordinary understanding of persons in this world. In hypothetical worlds where such disembodied entities might exist, they should not necessarily be considered persons.

2.4 Conclusion

This is a long way from Locke's original concept of personal identity, but advancements in science demand theoretical evolution. If the presence of a physical body is necessary as evidence for a person's continued existence but not sufficient, then it must be its combination with something else that grounds personal identity. And if the metaphysics of persons cannot rely solely on the mind, the latter can still plausibly serve as the condition of individuation. That is, if the body allows for the existence of a person at each moment in time, then the identity of that person can arguably be guaranteed by the continuity of the mind.

The body is the minimal condition necessary for personal identity to arise, while the mind is the fundamental condition: incapable of sustaining personal identity without the body's prior existence. This minimal condition is not merely a figure of speech; it must also account for the possibility that one body may host multiple streams of consciousness. As long as there is one body – one causal interface with the world – there may be multiple persons (e.g., individuals with DID, where distinct 'alters' may be regarded as different persons, in a way that remains compatible with persistence understood as perdurance). In this sense, the hybrid view of personal identity defended here still qualifies as neo-Lockean.

As from the beginning with Locke's account, the continuity of mind and consciousness still appears, in recent studies, to depend heavily on the continuity of memory. A neo-Lockean approach may still prevail by acknowledging the necessity of the body while maintaining the sufficiency of the mind. The next question will be this: if memory and mind are to remain fundamental criteria for personal identity, then we need a theory of mind that does not presuppose personal identity. Identifying such a candidate will be the main concern of the next chapter.

Part II: Inadequate Solutions & Their Limits

In this part, I critically examine the dominant causal theories of memory and their struggle to adequately explain the first-person perspective and immunity to error through misidentification. Chapter 3 focuses on the shift from causal to post-causal theories, addressing the inherent difficulties of grounding personal identity solely in causal continuity. This chapter highlights why purely causal accounts face significant philosophical challenges and why loosening causal conditions alone is insufficient.

Chapter 4 explores the first-person dimension of memory through the notion of quasi-memory and its relation to immunity to error through misidentification. By analyzing quasi-memory and related concepts, I argue that existing theories do not fully account for the epistemic status of memory-based 'I'-thoughts. Together, these chapters highlight the limitations of current approaches and lay the groundwork for more promising alternatives that emphasize the minimal necessity of bodily criteria and advocate for the simulation theory of memory, which will be developed in subsequent parts.

3 From Causal to Post-Causal Theories of Memory

Traditionally, memory has been seen as the means by which we know about our own personal past, because we naturally think that when we remember, we do so quite faithfully to what actually happened during a past experience. Most of the time, when we recall an episode from our past, we do not systematically doubt its accuracy. This intuition stems from a presupposition dating back to Plato, who compared our memorial capacity to impressions on a wax tablet. In the second half of the 20th century, this image of memory inspired Martin and Deutscher (1966) to propose a necessary and sufficient causal condition that functions just like Plato's wax tablet. This causal view of memory not only influenced other philosophers of memory but also made a significant impact in the literature on personal identity. Shoemaker (2003a, 2008) and Parfit (1984), for example, have written extensively on the relationship between the causal character of memory and the identity requirement inherent in that causal character, which locks memory-based personal identity into a problematic circularity.

However, the literature on memory has progressed significantly over the past 50 years, with philosophers increasingly considering memory to be less and less causal. The causal condition, once held as necessary and sufficient in Martin and Deutscher's (1966) claim, is now regarded by some as perhaps not even necessary (Michaelian & Robins, 2016). That said, within the personal identity literature, there has been comparatively little movement toward adopting non-causal theories of memory as the foundation for personal identity theories. Although some philosophers and psychologists have mentioned such possibilities, either explicitly or implicitly (e.g., Schechtman, 1996; Klein et al., 2002; Klein & Nichols, 2012; D'Argembeau, 2012; Michaelian, 2016).

In this chapter, I will provide an overview of the evolution of memory theories, from Martin and Deutscher (1966) to more recent claims. I will connect the notion of the memory trace, which characterizes the causal condition, with the psychological continuity condition that underpins neo-Lockean theories of personal identity. I will argue that there is no way to avoid the circularity objection if neo-Lockean theorists continue to adhere to a causal view of

memory. Causal theories fail to simultaneously account for both the authenticity of apparent memories and a non-circular account of identity.

3.1 Causal Theories of Memory

Theories of memory can be categorized in various ways. For example, De Brigard (2023) distinguishes memory as three different kinds of phenomena: an object of study distinct from imagination and perception; an action involving the retrieval of either a representation of a past experience or direct cognitive contact with that experience; and *something that is associated with certain entities*, such as objects, contents, vehicles, and traces. For the purposes of this chapter, I will not adopt this analytic framework. Instead, I will provide an overview of memory theories that best reflects the historical development – from the initial claim that memory is causal (marking the start of a significant literature stream) to more recent non-causal accounts that have emerged after decades of evolution.

The Classical Causal Theory. Taking memory representations for granted, Martin and Deutscher (1966) establish three conditions for someone to be genuinely remembering: (1) the person presently mentally represents the event, (2) the represented event was experienced by the rememberer at the time it happened, and (3) there is a causal connection between the person's current mental representation and the event experienced. They consider these conditions to be "separately necessary and jointly sufficient, if an event is to be an instance of remembering" (Martin & Deutscher, 1966: 166). However, condition (3) – the namesake of the theory – is what truly distinguishes it.

This causal condition boils down to a philosophical concept of the memory trace, and together, these two notions are supposed to differentiate remembering not only from imagining but also from relearning (Michaelian & Robins, 2018). If we once again take Plato's wax tablet as an example, memory traces for Martin and Deutscher are precisely that: there is "the existence of some sort of trace, or structural analogue of what was experienced" (Martin & Deutscher, 1966: 189). This "structural analogue" lies at the heart of the causal claim: if a person genuinely remembers, then there must exist within the person a memory trace of the remembered event – acting as a structural analogue between the actual past experience and its current mental representation. Although not stated explicitly, the theory implies that the person must be numerically the same at both t_1 , when the experience occurred, and t_2 , when it is remembered, in order for the trace to persist. It is in this way –

3 From Causal to Post-Causal Theories of Memory

through reliance on a causal theory that requires a memory trace – that neo-Lockean theories become vulnerable to the circularity objection.

Neoclassical Causal Theories. Several approaches to memory can be categorized as neoclassical causal theories. While they retain the core claim that a memory trace constitutes the causal connection between past experience and present representation, they introduce important refinements. These include: redefinitions of the nature of memory causation, variations in the degree of similarity required between the original experience and its representation, differing views on whether memory traces are localized or distributed in the brain, and a methodological shift toward more naturalistic and empirical approaches rather than purely philosophical intuitions.

For some theorists (e.g., Bernecker, 2008, 2010), memory causation does not depend on a single, specific memory trace tied to each remembered instance. Rather, one learns during the original experience and remembers through a network of traces. Memory causation, in this view, consists of multiple traces forming an uninterrupted chain between learning and remembering (Michaelian & Robins, 2018). Bernecker (2010: 217) also argues that the past experience and the present representation need not be identical, but only “sufficiently similar”. This claim is especially compatible with the distributed view of memory traces, which supports the idea of a continuous learning-remembering chain. Moreover, neoclassical causal theorists have developed strategies to address philosophical concerns about the concept of a memory trace – particularly the worry that it imposes an unrealistic burden on scientists to locate such a specific entity in the human brain. In response, philosophers have adopted more naturalistic approaches and collaborated more closely with empirical research, such as studies on the role of specific brain structures, primarily the hippocampus, in remembering (Michaelian & Robins, 2018).

Although these moves seem to loosen the original causal condition, the claims still fundamentally support a necessary causal condition, which, together with some additional criteria, is considered sufficient. They mark a significant shift from the original causal theory by no longer requiring the content transmitted from the past experience to be identical to the present representation; the neoclassical view allows for certain kinds of content modification during transmission. For example, they accept it as genuine remembering if I recall having received a book for my ninth birthday, even though I specifically received *Little Women*. However, since neoclassical causalists still adhere to the necessity of memory traces, their

view does not allow for content generation in genuine remembering. That is, the remembered content may be *less* detailed than the original experience, but it cannot be *more* informative.

Hybrid Theories. We can characterize hybrid theorists as those who agree that the causal condition is necessary but deny that it is sufficient (Michaelian & Robins, 2018: 19). Proponents of hybrid theories tend to introduce an additional criterion alongside the causal condition. For example, Martin and Deutscher's oft-cited painter case – in which a painter depicts a scene from his own past experience without consciously recognizing it – has prompted discussions around the criterion of epistemic relevance (Debus, 2010; Fernández, 2019).

From the line of discussion on the phenomenology of memory – or the feeling of remembering – some have argued for the necessity of including auto-noesis, or auto-noetic consciousness, as a required condition (Klein, 2014; Klein & Nichols, 2012; Dokic, 2014). According to Tulving (1985), who distinguishes three levels of consciousness – auto-noetic, noetic, and auto-noetic – these correspond respectively to non-declarative procedural memory, semantic memory, and episodic memory. It is episodic memory, which involves auto-noetic consciousness or self-awareness, that interests philosophers due to its personal nature and its special capacity to provide individuals with knowledge about their own past. With cases such as the painter case, and real-life examples like that of patient R.B., it appears possible to access one's personal past experience without experiencing a sense of ownership over it. This is why hybrid theorists argue for the inclusion of the auto-noesis criterion and do not consider the painter case or patient R.B.'s case as instances of genuine remembering.

Although hybrid theories seek additional conditions to complement the causal theory, the causal condition itself remains unchanged. Some argue that, for this reason, hybrid theories are simply variants of causalism, since these additional criteria can be easily adopted by causal theorists as part of a jointly sufficient set of conditions – without altering the role of memory causation (Michaelian & Robins, 2018).

3.2 Causalism and Circularity

The circularity objection was originally raised against Locke by Butler, who argued that the memory criterion of personal identity is viciously circular.¹² He pointed out that memory

¹² In neo-Lockean theories, the memory criterion has evolved into the psychological continuity criterion. I take it for granted that anything that applies to memory also applies to other mental states.

3 From Causal to Post-Causal Theories of Memory

cannot define personal identity “because memory already presupposes personal identity – as knowledge in general presupposes truth” (Noonan, 2019: 11). To survive this circularity objection, there are two possibilities: either we accept that memory is not a necessary condition for personal identity, or we reject Butler’s claim that memory presupposes personal identity.

Undertaking the first strategy would mean endorsing strong animalism and accepting a number of unattractive assumptions that come with the view, such as: that the embryo I once was and the person I am today are the same despite the psychological discontinuity between the two; and that a person in a vegetative state is the same as the lively individual they were before the tragic accident. Although some do advocate for this first possibility (e.g., Snowdon, 2014; Olson, 2002, 2003), in order to avoid committing to such a radical position and to remain faithful to the neo-Lockean spirit of this work, I will be exploring the second possibility.

To object to Butler’s circularity claim, it is crucial to clarify why we intuitively think that memory presupposes personal identity – that is, what exactly Butler had in mind when he argued that memory presupposes identity. In standard presentations of the causal theory of memory, the subject “*S*” is assumed to be the same person at two times: *P1* at *t1*, who experienced the event, and *P2* at *t2*, who remembers it. The causal connection is then understood as linking the same person across time. Without assuming this identity, the causal relation itself would not properly connect a subject to their own past experience. In possible worlds where this connection is missing, such as in cases of quasi-memory, *P2*’s remembering an event experienced by *P1* does not entail that *P2* is *P1*. Thus, the causal condition presupposes the identity of the subject at *t1* and *t2* to establish an appropriate causal link, rather than establishing identity itself. According to causal theories, if there is a memory trace linking *P2*’s present representation to the past event, we take this as evidence that *P2* and *P1* are the same person, and that the remembering is genuine.

So, the discussion over the nature of the memory trace can actually be considered a discussion about the nature of the identity relation between *P2* and *P1*. Neoclassical causalists suggest that the memory trace shouldn’t be conceptually considered as one single continuous trace that is fixed, with explicit contents and carried by one distinct local vehicle; rather, it should be understood as a network of distributed vehicles – complex entities stored in different locations. Each of these locations provides locally stored memory content, each

carried by a distinct vehicle (Michaelian & Robins, 2018: 21). In this way, the concept of memory remains causal since the concept of a memory trace is still valid. Therefore, P2 is identical to P1 in the sense that the active memory trace providing P2 with the apparent memory is a network of distributed vehicles of memory contents. This means memory is “reconstructed” from the distributed vehicles rather than “reproduced” from one single local vehicle considered as the memory trace (Sutton, 1998: 2; Michaelian & Robins, 2018: 21).

Now, it seems that between reproduction and reconstruction, the distinction primarily concerns the nature of the mnemonic representation (the memory content), rather than the cross-temporal identity link secured by the memory trace. Neoclassical causalism allows for reconstruction in memory content, which introduces the notion of degrees of *similarity* into memory causation, replacing the earlier demand for strict *identity* of memory content. For example, I can remember quite precisely that I received a copy of *Little Women* as a present for my ninth birthday; I can remember more generally that I received a book as a birthday present when I was little; and I can also remember vaguely that my parents used to throw me birthday parties where I received presents. The notion of memory remains causal and compatible with reconstruction, meaning that it does not require identical memory content for an instance to be considered genuine remembering. The nature of the causal condition is thus modified so that the personal identity criterion no longer relies on *strict identity* of memory content but rather on *sufficient similarity*.

The notion that memory does not reproduce an identical representation of the past experience is consistent with weak preservationism. Were memory to require strict identity between the experiential content and the retrieved representation, this would constitute strong preservationism, a position that appears problematic due to its incompatibility with the natural occurrence of forgetting. Nevertheless, weak preservationism still fundamentally depends on the causal reliability of the distributed memory vehicles in forming the memory trace. Specifically, a causal relation between *P1* and *P2* – characterized by personal identity – is necessary: the activation of these vehicles to generate an apparent memory for *P2* presupposes that the same vehicles were originally instantiated during *P1*'s experience of the event. Thus, the relaxation of the causal condition from strict identity to content similarity does not dispense with the requirement of personal identity. Put differently, any account of memory permitting similarity in content remains contingent upon a causal connection that secures the identity of *P2* and *P1* as the same person.

If neither strong nor weak preservationism is compatible with an identity-less memory theory, can generationism, which allows for non-causal reconstruction in memory, make a stronger case? Let us now turn to consider whether post-causal theories offer a promising alternative in this regard.

3.3 Post-Causal Theories

The primary challenge facing causal theories lies in their incompatibility with memory's reconstructive and generative character. Although few today would endorse strong preservationism of memory content – that is, the claim that the retrieved representation and the original experiential content must be exactly identical – weak preservationism remains widely accepted. This view holds that while the two need not be identical, the retrieved representation must at least be part of, or logically implied by, the content of the original experience (Michaelian & Robins, 2018). In other words, the causal condition is generally regarded as necessary but not sufficient on its own to fully define genuine remembering. Post-causalist theories go further, arguing not only that the causal condition is insufficient, but also that it is not necessary.

If causal theories treat memory as a *diachronic* capacity, post-causal theories treat it as a *synchronic* capacity, in which the occurrence of remembering depends on “what happens when the subject (apparently) remembers, rather than on whether there is a suitable relationship between the subject's retrieved representation and his experiential representation” (Michaelian & Robins, 2018: 24). Loyal to this general idea, several post-causal theories embrace the reconstructive character of memory in a generative way that causalism cannot.

First, there is the narrative theory of memory (Schechtman, 1996, 2011b), which is not exactly a full-blown memory theory but rather a personal identity theory based on what should count as genuine memory so that one's personal story makes sense to oneself. However, it is possible to deduce a memory theory from the narrative theory of identity. Instead of saying that we make up stories of our past to create our own identities, the view holds that “in memory, we build stories about our past; stories which include content that we have acquired through our own experience and content from other sources, like imagination or testimony” (Fernández, 2019: 41). Rather than being completely loyal to the objective past, memory is the reconstructed retrieval of the objective past, which provides us with a greater

sense of the narrativity of our own lives than what might have actually happened. Schechtman (1996, 2011b) developed this view because it sheds more light “on the relation between personal identity and survival, moral responsibility, self-interested concern, and compensation” (1996: 3). We might say that, according to the narrative theory, a person is constituted by the version of the past that makes sense to them, and this version of the past is represented by a reconstructed apparent memory that can be either very different from the original experiential content or about something that might not have even happened to the subject.

This is compatible with generationism in the sense that, as long as the apparent mental representation makes sense and feels like an instance of memory to the individual, it does not need to be causally connected to the original experience to be considered genuine remembering. For example, patients with Korsakoff’s syndrome – a form of amnesia typically caused by long-term heavy alcohol consumption – can provide detailed descriptions of events that never actually happened to them. These descriptions are purely generated, stemming from imagination or what might be called daydreaming. However, according to the narrative theory, such instances still count as legitimate memories as long as they make sense to the subject within the narrative framework of their personal past.

Second, there is the functionalist theory of memory. Fernández (2018, 2019) proposes an account of memory based on the nature of its functionality: as long as an instance functions for a subject as an episode of memory, it can be qualified as a memory instance. Functionalism is designed to incorporate, first, the etiology and self-reference of the mental state as its heritage from causalism, and second, the impact the mental state has on the subject’s beliefs. Fernández formalizes this as follows: “for any subject *S* and event *e*, *S* remembers *e* just in case (1) *S* has some mental image *i* such that *i* ‘tends’ to cause in *S* a disposition to believe both that *e* happened and that *S* experienced *e* to happen, and (2) *i* ‘tends’ to be caused in *S* by having experienced *e* to happen” (Fernández, 2018: 64, numbering added). The first part of this formalization, showing that memory tends to cause belief, is introduced as the ‘mnemonic role’ of memory; the second part shows that memory tends to be caused by past experience. The departure of functionalism from causalism lies in the idea of tendency: “it does not require that the mental state actually be caused by the experience” (Michaelian & Robins, 2018: 24).

3 From Causal to Post-Causal Theories of Memory

According to functionalism, as long as the mental state plays the mnemonic role in the subject's psychology, then regardless of whether the mental state is actually caused by a personal past experience, it qualifies as memory. For example, a patient with Korsakoff's syndrome would not be considered to be remembering according to functionalism. First, regarding the mnemonic role: since it is known that the patient suffers from amnesia, if the patient had truly experienced the event as described, they would not have the same type of mental images as those presently described. Thus, the mental state is not playing the mnemonic role that corresponds to the patient's psychological environment. Second, regarding the tendency of causation: if the patient's present mental representations are not of the kind that tend to be produced by past experiences, then, according to functionalism, the current mental states do not qualify as instances of memory (Fernández, 2018: 66). Fernández argues that functionalism is compatible with generationism, as long as the generation of the mental representation does not undermine these two conditions.

Finally, there is the simulationist theory of memory. The simulation theory was first developed as a theory of mind (Shanton & Goldman, 2010) and later discussed in the philosophy of memory (De Brigard, 2014; Michaelian, 2016). Unlike functionalism, simulationism involves an investigation of the role of the memory trace, which can be considered its heritage from causalism. Before developing non-causal simulationism, Michaelian (2011) first discussed a constructive causalism that both requires the transmission of content and permits the generation of new content. However, he points out in subsequent work (2016) that "the very same cognitive process may be at work both in cases in which some content is transmitted from the experience and in cases in which no content is transmitted" (Michaelian & Robins, 2018: 26). Through a naturalistic approach supported by empirical evidence, simulationism claims that remembering is not different from imagining, in the sense that the same part of our brain, namely the hippocampus, is active when the subject is either remembering or imagining (Michaelian, 2016: 99–102). In fact, it is empirically "impossible to draw a firm distinction between remembering and imagining the past" (Michaelian, 2016: 100).

Simulationism is compatible with generationism in the most direct way: it allows as much generation of new content in remembering as in imagining.¹³ This account provides a symmetry between remembering and imagining: just as there should be a limit to the episodic constructive process beyond which an instance is disqualified as remembering (which would then be an instance of misremembering one's past), there should also be a limit to future-oriented imagining beyond which an instance is disqualified as episodic imagining (which would be an instance of misimagining one's future). For example, I can remember very little about my earliest birthdays, but I might have heard about some of them while growing up. Say, there was this one specific birthday party – my fourth birthday – that was especially memorable among my parents and other relatives because I received a piano from my father and knocked over my juice on the leather cushion of the brand-new piano chair. Growing up hearing about this event, I have reconstructed several different versions of this episode. Now, when I remember the scene, I can effortlessly mentally represent the piano because it accompanied me throughout my youth, and I can also picture how the juice – of some fruit, of a certain color – flew out of its plastic cup, also of a certain color. This instance qualifies as simulationist memory, but would never be qualified as any kind of causal memory.

3.4 Post-Causalism and Circularity?

What insights does this overview of post-causal theories provide regarding the problem of circularity in neo-Lockean theories of personal identity?

According to the narrative theory, as developed by Schechtman (1996, 2011b), personal identity is grounded in the capacity to construct a coherent and intelligible narrative of one's life. Within this framework, memory plays a crucial role – not because it preserves an objective past, but because it contributes to a self-narrative that is intelligible to the person. However, this reliance on narrative intelligibility raises concerns, since what counts as “making sense” is inherently subjective and shaped by the individual's internal psychological context. Narrativism, in this more precise form, can account for memory-related disruptions in cases such as dementia or amnesia, where individuals struggle to maintain a coherent self-narrative (Schechtman, 1996: 146–147). These individuals are not simply missing facts about their past;

¹³ Questions regarding the limits of simulationist memory in terms of epistemic power and the overall reliability of memory will be discussed in the final chapter of this thesis. Here, I would like to focus on how post-causalist theories are non-causal.

3 From Causal to Post-Causal Theories of Memory

they are losing the structure that allows them to understand themselves as the same person over time. However, it remains challenging to determine how narrativism can adequately address cases of confabulation, where fabricated memories might also contribute to a seemingly coherent narrative, despite lacking factual grounding.

It seems that, to endorse narrativism, one must accept that patients who constantly confabulate are genuinely remembering – even when the mental representations they claim as memories are improbable or even impossible. Narrativism would need to provide a framework that treats such instances as genuine memory, which appears highly challenging. For example, patients with schizophrenia spectrum disorders are known to make more source-monitoring errors during recollection (Mammarella et al., 2010). Although their level of imagination inflation is similar to that of healthy individuals, they are more likely to mistake imagined representations for personal memories, suggesting a failure or deficiency in their metacognitive monitoring of memory sources. Nonetheless, by taking these representations as memories, these patients are able to construct a narrative of their personal past that remains perfectly coherent to them.

It seems difficult, in such cases, to attribute the special epistemic power of episodic memory to narrativist memory. Narrativism is compatible with generationism, but it is not compatible with the mnemonic epistemic function of identifying confabulation. Therefore, even though the aim of the present research is to move beyond the causal condition of memory, the narrativist approach appears too loose – borrowing Fernández's (2019) term – in the sense that it accepts far more cases as memory than it arguably should.

Functionalism requires genuine remembering only to have the tendency of reliably representing one's personal past, treating the generation of memory content as occasional and partial – acceptable only if the representation is of the right kind functionally speaking, namely, if it plays the mnemonic role (Fernández, 2018, 2019). The expression tendency seems to suggest, on the one hand, a general alignment between experiential and retrieved content in episodic memory, and on the other, the acceptance of occasional generation of memory content. The challenge faced by functionalism, then, is to provide evidence that the generation of memory content is indeed only *occasional* and partial, rather than *habitual* and *widespread* in human memory.

If, by *tendency of representing one's personal past*, we mean that a certain degree of memory generation is included, then generation would not be merely occasional but habitual;

therefore, the distinction between representations of the right kind and occasional generation should not exist. If, on the other hand, it refers to a form of genuine remembering that preserves rather than generates content, then functionalism appears to be stricter than expected. Although the concept of the functional mnemonic role of memory seems relatively clear, it appears that, as long as the expression tendency remains ambiguous, it will be difficult to evaluate functionalism's compatibility with generationism. Furthermore, Fernández (2019) emphasizes the self-referring character of memory as reinforcing the idea of *tendency*. However, if the expression is interpreted in the second sense, functionalism risks falling back into a form of causalism.

Simulationism is designed to fully endorse the generation of memory content. It considers remembering as personal past-oriented imagination, which is symmetrical to personal future-oriented imagination – the process that gave rise to the expression “mental time travel”. However, being fully non-causal, it faces challenges concerning the epistemic power of episodic memory. If remembering is a kind of imagining, what becomes of memory knowledge? In response to this fundamental challenge, Michaelian (2016) provides evidence from multiple directions to argue that simulationist memory remains generally reliable for memory knowledge, despite being considered a kind of imagining. Just as we are able to imagine things that are more likely to happen to us personally in the future (e.g., I will bike to campus tomorrow) and things that are only logically possible in other worlds (e.g., I will fly to campus on the back of my dragon tomorrow), we similarly do the same with remembering: I can remember things that closely resemble what actually happened in my past experience (e.g., I biked to campus yesterday), and I can also claim to remember things that are highly implausible or even impossible (e.g., I flew to campus on the back of my dragon yesterday). This distinction corresponds to that between episodic and pure imagining, and between episodic remembering and misremembering.

Therefore, the main challenge for simulationism seems to be how it should delimit what is episodic – or what is about our personal story – and what is not. However, this challenge is not about distinguishing one's own episodic imagining and remembering from someone else's, but rather distinguishing one's episodic imagining and remembering from misimagining, misremembering, or counterfactual thoughts. This challenge for simulationism, then, highlights one of its most important virtues: it provides a definition of memory that doesn't require personal identity as a criterion. Simulationist memory is capable of solving

3 From Causal to Post-Causal Theories of Memory

classical puzzle cases of personal identity, such as: If I don't remember being born, am I the same person as the baby who was born? If memory-based personal identity theories are correct, and we adopt simulationist memory, we can easily say that I do remember episodically through testimony from my parents and biological knowledge of childbirth, and therefore, I am the same person as the newborn baby.

It seems that simulationism is capable of endorsing both narrativism and functionalism, but the compatibility does not work in the opposite direction. That is, simulationism can account for cases of constructive narrativity while excluding confabulations. Narrativism, on the other hand, includes cases that simulationism would not count as episodic remembering, accepting them as genuine memory for the sake of their coherence within a personal narrative, even when they involve confabulation. For example, occurrent representations in patients with schizophrenia would not qualify as simulationist memory due to source-monitoring errors. Simulationism can also account for cases recognized by functionalism, as well as those left out by the stricter criteria of functionalist memory. For instance, in the childbirth puzzle case, my generated and reconstructed memory of being born might not clearly count as functionalist memory if one interprets the functionalist 'tendency' as requiring a closer link to past experiences. However, if we understand 'tendency' in a looser sense – as allowing for some reconstruction, as long as the representation generally plays the mnemonic role – then functionalism might also be able to accommodate such cases. This nuance highlights that simulationism may offer a broader but still epistemically responsible framework for defining memory.

In this sense, simulationism appears to be the most promising of the post-causal theories for offering an identity-less definition of memory – one that might ground personal identity without falling into circularity. This raises a crucial question: how can a definition of remembering that does not presuppose personal identity nonetheless place sufficient constraints to support it? Importantly, simulationism is grounded in empirical findings about the brain's metacognitive system, particularly the hippocampus, which plays a central role in episodic remembering. However, while these findings show the neurological basis of simulationist memory, they do not imply that bodily continuity or biological conditions are necessary components of simulationism itself. Rather, if one wishes to incorporate simulationism into a neo-Lockean framework for personal identity, a hybrid account may be required – one in which memory and psychological continuity remain central, but where a

minimal bodily criterion is invoked to address the explanatory gaps left by purely psychological theories.

3.5 Conclusion

In this chapter, I have traced the evolution of memory theories from causal to post-causal approaches, moving from a nearly strong preservationist stance to endorsing generationism. I have shown that although the causal condition in causalism does not directly imply personal identity, the identity criterion is necessary for memory traces to fulfill the role that causal theorists attribute to them. I have discussed the shortcomings of classical, neo-classical, and hybrid causal theories, emphasizing that despite advances in other areas, the fundamental causal condition remains unchanged. Turning to post-causal theories, I have demonstrated how simulationism can account for narrativity in a stricter manner and functionalism in a more permissive way. Finally, I identified the potential for a simulationist memory-based theory of personal identity, suggesting the likely necessity of incorporating some minimal bodily criterion.

This conclusion opens the way for a deeper investigation of personal identity beyond purely memory-based accounts, as simulationism is not the first proposed solution to the circularity problem. Reluctant to abandon completely a causal condition, Shoemaker developed the notion of quasi-memory and endorsed its weaker version. He argues that, in light of the nature of quasi-memory, memory can constitute personal identity without falling into circularity. In Chapter 4, I turn to the analysis of the self as a linguistic object, drawing on Anscombe's influential work on agency and action. Following her argument that the use of 'I' does not designate a particular object or entity, I explore how this perspective challenges intuitive notions of personal identity. Applying this framework to memory claims reveals an inherent instability in personal identity, which I illustrate through the concept of quasi-remembering in possible worlds where such phenomena are both possible and commonplace.

4 The First Person, IEM, and Quasi-Memory

The special nature of first-person thought has triggered a rich stream of literature since its formalized introduction into modern philosophy through Descartes's infallibility thesis in the *Meditations*, which defends the idea that "we cannot be mistaken about our current conscious states" (Coliva & Palmira, 2025). In the second half of the 20th century, Wittgenstein, in his *Blue Book* (1958), further specified that the distinctiveness of first-person thought lies not in a general infallibility but in the "impossibility of misidentifying the subject of these thoughts" (Coliva & Palmira, 2025). Shoemaker (2003b) later formalized this notion as immunity to error through misidentification (IEM). Anscombe (1975) then deepened the investigation into the function of 'I', arguing that it is not a referential term. This view has sparked multidisciplinary discussions across epistemology, the philosophy of language, the philosophy of mind, and metaphysics, particularly concerning memory, introspection, self-knowledge, self-reference, and self-consciousness (Coliva & Palmira, 2025). Memory-based self-ascriptions became central to the Shoemaker–Evans debate, which led to a categorization of different types of first-person judgments, various forms of IEM, and competing accounts explaining why certain classes of judgments are IEM.¹⁴

Philosophers working on IEM primarily position themselves by debating whether the use of 'I' in memory claims is, in some sense, IEM. For example, Coliva and Palmira (2024) defend a Shoemakerian line by arguing that introspection-based self-ascriptions of psychological properties are the only cases of logically IEM self-ascriptions (Coliva & Palmira, 2025: 5), and that a metasemantic account of IEM is preferable to other explanations. Pryor (1999) also analyzes Evans's arguments before ultimately siding with Shoemaker in maintaining that memory-based uses of 'I' exhibit only *de facto* IEM.

This brief overview demonstrates how, over time, the discussion of the first person and its special status has influenced the literature on memory. That specialness highlights not only the 'subjective character' of the first person but also raises questions about the ontology

¹⁴ For a detailed and comprehensive account of these categories, see Coliva and Palmira (2025).

of memory itself. In this chapter, I examine Anscombe's analysis of 'I' as a non-referential term, drawing on insights from interpretations of her philosophy of action. Anscombe's account of a non-referential, object-less first person will then be tested through the concept of quasi-memory to assess whether such a definition of 'I' can ensure IEM in other possible worlds. I will conclude that it cannot, and that it does not seem necessary to preserve logical IEM if we want to move beyond causal theories of memory in defining personal identity – since in possible worlds where quasi-remembering does not concern our own personal past, identity does not appear to matter in the same way it does in our world.

Let us begin with the insights provided by Anscombe's view on 'I'.

4.1 'I'-Thoughts and Non-Observational Knowledge

In her essay "The First Person", Anscombe challenges the common view that 'I' is a referential term that denotes the self as an object. She begins her argument by comparing 'I' to other categories of referential terms, such as names and descriptions. She imagines a peculiar society "in which everyone is labelled with two names": one visible to others and the other a private "A", stamped on the wrist of each individual (Anscombe, 1975: 49). In this society, each person refers to themselves as "A" instead of using 'I', as we do in our world. What strikes us as strange about these "A"-users is that, even though the use of "A" is private in the sense that each individual knows whom they are referring to, their reports of their own actions are given "straight off from observation" (Wiseman, 2017: 543), rather than in the immediate, first-personal way typical of our use of 'I'. When we use 'I', there seems to be a kind of immediacy that is incompatible with statements based on observation. Of course, there are many potential counterexamples to this claim – unless we first introduce Anscombe's selection of special use cases of 'I' that demonstrate how it does not function like a referring term.

To distinguish her position from the Cartesian view that the ego is determined through first-person thought, Anscombe identifies a class of first-person thoughts that are externally verifiable. These include not only actions (which indeed make up a significant portion) but also thoughts involving intention. Actions are observable by others through perception, while intentions are verifiable through the answers to a series of "why" questions. The distinction between the use of "A" and 'I' – as linguistic practices – is marked by this class of first-person

4 The First Person, IEM, and Quasi-Memory

thoughts. This “linguistic practice to which ‘I’ belongs is connected to the linguistic practice described in Intention” (Wiseman, 2017: 543).

To see the deep connection between the reference-less ‘I’ and Anscombe’s philosophy of action, it is necessary to understand her concepts of ‘knowledge without observation’ and ‘practical knowledge’. According to Anscombe, these two types of knowledge help justify the non-referential nature of ‘I’. Knowledge without observation can be expressed through the class of first-person thoughts defined by Anscombe: those concerning actions and intentions (Anscombe, 2000). This is a kind of knowledge an agent has about their own actions, which does not require observation in order to be epistemically justified.

It is easier to grasp what Anscombe means by considering non-observational knowledge concerning actions. Actions involve interactions between our bodies and the physical world, and in ordinary situations, we know about our own bodies without needing to observe them. For example, I (usually) do not need to look at my leg to say, “I have my leg at this spot”. Excluding severe clinical cases such as vegetative states, or temporary situations like partial paralysis caused by sitting too long in one position, we typically have knowledge of our bodies without needing to see them. This special capacity to say (Wiseman, 2017: 546) something about one’s own body without observation, for Anscombe, illustrates the entry level of the specialness of ‘I’.

Non-observational knowledge doesn’t necessarily have to concern bodily movements; first-person thought can also pertain to intentional knowledge. This brings us to the heart of Anscombe’s philosophy of action, where she distinguishes between cause and reason in explaining actions: “a mention of something past gives a cause rather than a reason for what I did” (Wiseman, 2017: 544). When a character in a Greek tragedy says, “I killed him because he killed my father”, there is a clear, identifiable reason for the killing. By contrast, when I say, “I jumped because the coming car honked”, the honk is not a reason but a cause of my action. This comparison marks the beginning of Anscombe’s discussion on practical knowledge and practical reasoning. Practical knowledge is the knowledge contained in the agent’s intention: it aligns with the reasons and objectives behind a series of intentional actions.

When something in the past is the reason I do something, I have reasoned practically to “do what happens” (Anscombe, 2000: 52), and this brings out what intention means to Anscombe. This practical knowledge, like the non-observational knowledge of bodily actions, is characterized by immediacy and directness. Anscombe considers intention a kind of non-

observational knowledge and compares it to bodily movement; she describes intentions (as practical knowledge) as “special interior movements” of the first person (Anscombe, 2000: 49).

Suppose the Greek character hasn’t yet killed the man, but has the intention of doing so: he does not need to observe himself introspectively to know that he is on his way to kill the man who killed his father. This intentional knowledge is immediate and direct – it needs no further justification. Anscombe illustrates this point with an example of a man who is assassinating an entire community of evil politicians by plumbing poisonous water into their households. The man is (A) moving his arm up and down to (B) plumb the water, to (C) eliminate them all, in order to (D) “institute the Kingdom of Heaven on Earth” (Anscombe, 2000: 37). If we ask the man what he is doing and why, and he responds with A, B, and C, we can uncover D – not a bodily movement, but an intention being realized through a sequence of physical actions. In Anscombe’s sense, the man is reasoning practically. Practical reasoning enables us to uncover intentional knowledge, which is non-observational for the subject.

Anscombe’s view is known to be quite counterintuitive. If she claims that ‘I’ does not refer to an object – a metaphysical self – then what is so special about ‘I’? When the plumbing man says “I’m eliminating them all”, if there is no thinking thing (in Locke’s terms) occupying the slot of ‘I’, then what is doing the practical reasoning?

The answer to the first question lies in the linguistic structure. By rejecting the referential view, Anscombe is not leaving us with ‘I’ as a meaningless “linguistic device”. Instead, she replaces the “picture of ‘I’-use as involving a simple capacity (reference) with a complex object (self)” with “a set of complex capacities the exercise of which is not mediated by sensation and so does not require picking out any object”. Moreover, the fact that these capacities yield verifiable truth-values “underpins the ‘logician’s rule’: If X asserts something with ‘I’ as subject, his assertion will be true if and only if what he asserts is true of X” (Wiseman, 2017: 550; Anscombe, 1975: 60).

What is special about ‘I’ is not that it refers to a special object (the self), but that it plays a special role in our linguistic and epistemic practices – as a capacity for self-ascription. As Anscombe points out, our tendency to fall into the trap of seeking a referent for ‘I’ stems primarily from the long-standing influence of Cartesian dualism.

The second question also reveals our intuitive urge to presume the existence of a preexisting metaphysical self: a kind of Cartesian ego that is independent of and prior to bodily

movements. Anscombe explicitly denies this notion. This rejection is one of the most fundamental reasons her view has remained highly controversial since its publication (Wiseman, 2017: 537). However, this does not mean that she denies the existence of self-consciousness. In her discussion of intention, she describes the special “capacity to say” characteristic of ‘I’-users as self-consciousness (Wiseman, 2017: 550). This invites the interpretation that, although Anscombe rejects the view of ‘I’ as a referential term, she still affirms that the first person’s capacity to self-ascribe is grounded in a form of self-consciousness. Crucially, however, this self-consciousness does not identify itself with a particular object or metaphysical entity, as she makes clear in her 1975 essay when criticizing Cartesian assumptions.

Anscombe’s view on the first person opens up the possibility of a concept of ‘I’-thought – including memory-based ones – that is independent of reference and identification. This is particularly attractive for a project seeking a potentially non-circular neo-Lockean theory of personal identity. However, Anscombe’s claim regarding the absence of an object seems to rely heavily on the presence of bodily movements and sensations. In the context of this research, the discussion of the first person and reference raises further concerns, particularly in relation to memory-based self-ascriptions. One is the concept of quasi-memory, developed to loosen the causal condition of standard memory in order to avoid circularity in defining persons through memory. The other is IEM, which highlights the seemingly special epistemic status of ‘I’-thoughts.

In the next two sections, I will first turn to memory-based ‘I’-thought and the concept of quasi-memory, in order to examine whether the attempt to slightly loosen the causal condition is sufficient to overcome the circularity objection, and to explore how an object-less concept of ‘I’ might align with quasi-memory. Then, I will move on to discuss IEM in more detail.

4.2 Memory-based ‘I’-Thoughts and Quasi-Memory

The reason Shoemaker develops the concept of quasi-memory is that, if Anscombe is correct in claiming that “I am E.A.” does not refer to a metaphysical self and “is after all not an identity proposition” (Anscombe, 1975: 60), then the coherence of self-reference in memory-based judgments is undermined. Memory claims such as “I was in Rome last week” lose much of their meaning, because ‘I’ no longer refers to any particular object. This brings us into the

realm of discussion about what makes memory-based 'I'-thought special, namely, the question of what constitutes genuine remembering, and what in genuine remembering provides persons with "a special access to facts about their own past histories and their own identities, a kind of access they do not have to the histories and identities of other persons and other things" (Shoemaker, 2003a: 19).

Compared to Anscombe's analysis of the nature of 'I', memory-based 'I'-thoughts complicate the immunity of 'I'-uses to EM, especially because the additional temporal dimension introduces specific criteria necessary to distinguish between different types of IEM.¹⁵ According to Anscombe, the use of 'I' itself functionally justifies its immunity: the absence of a metaphysical self makes the linguistic use of the first person (the capacity to say) straightforwardly, immediately, and directly immune to EM. If the statement "I am eliminating them all" contains non-observational knowledge of the speaker's intention to bring about the Kingdom of Heaven on Earth, then the statement "I was in Rome last week" contains memory-based knowledge of the speaker's past experience of being in Rome. However, unlike the IEM of the first kind – which is atemporal or ongoing – it has been argued by Nerlich (1959: 208) that, considering all possible worlds, "it is logically possible that we should have knowledge of past events which we did not ourselves witness, of experiences we did not ourselves have, and of actions we did not ourselves perform" (Shoemaker, 2003a: 23). In other words, it is logically possible to genuinely remember things that one did not actually experience first-hand, which challenges the identity criterion for genuine remembering. To explore this possibility, Shoemaker (2003a: 24) introduces the concepts of 'quasi-remembering', 'quasi-memory', and 'quasi-memory knowledge'.

In support of a causal view of memory, Shoemaker (2003a: 19–20) identifies two necessary conditions for it to be true that a person remembers a past experience: (1) the previous awareness condition and (2) the condition of IEM. In certain possible worlds (such as

¹⁵ All of Anscombe's examples of 'I'-use are essentially atemporal – that is, the presence or absence of a temporal dimension does not significantly alter their structure or meaning. In contrast, memory-based 'I'-thoughts introduce temporality into the very structure of 'I', since they involve a relation between the present and a past cognitive state. For example, in the statement "I remember I was in Rome last week", the first 'I' refers to the speaker in the present, while the second refers to the same subject at a time in the past. This is due to a correspondence between the current cognitive state and a past cognitive or sensory state involved in the original experience (Shoemaker, 2003a: 23).

The notion of IEM is typically developed in discussions of memory, so it might seem odd to apply it to atemporal 'I'-thoughts. However, in this context, I use IEM more generally to refer to the functional specialness of 'I': that is, its inherent resistance to misidentification whenever it is properly used. On this view, the specialness of 'I' in Anscombe's atemporal examples is functionally equivalent to IEM.

4 The First Person, IEM, and Quasi-Memory

those described by Nerlich), quasi-remembering differs from genuine remembering, generating a type of memory in which the genuine memory familiar to our world is only a special case. The weakened previous awareness condition marks the key distinction between the two. In quasi-memory, the quasi-rememberer need not have been present, cognitively or sensorily, at the spatiotemporal location of the remembered event; it suffices that they acquire quasi-memory knowledge after the fact, based on available data. In such cases, the spatial criterion is also largely irrelevant. It appears, then, that in these possible worlds, an identity criterion is needed to distinguish genuine or veridical memory: i.e., memory that truly concerns one's own personal past.

Quasi-remembering can be further distinguished by two levels of strength with respect to the previous awareness condition. A more radical version of quasi-memory is entirely free of causal constraints, where the previous awareness condition involves no spatiotemporal continuity in the causal chain. The second version – favored by Shoemaker (2003a) – does require a spatiotemporally continuous causal chain (albeit a much weaker version of the condition compared to that in genuine, veridical remembering) to serve as the identity criterion. This causal requirement helps avoid situations in which quasi-memory claims provide no meaningful information either to the quasi-rememberer or to others, due to the possibility of massive reproduction in those possible worlds (Shoemaker, 2003a: 31; Strawson, 1959: 20). It thereby preserves the epistemic usefulness of quasi-memory in identifying persons. In the more radical version, however, it becomes evident that the use of 'I' no longer possesses absolute or logical IEM. The concept of quasi-memory, especially in its stronger form, allows us to reflect on how 'I'-thoughts may operate in memory claims without presupposing a self as a particular object: i.e., without assuming the identity of any particular entity.

Favoring the weaker form of quasi-memory, Shoemaker (2003a) calls this kind of quasi-remembering “weak remembering”, while the normal memory in our world he terms “strong remembering”. The entire conceptual maneuver of creating and analyzing quasi-memory aims to overcome the circularity objection to Locke's personal identity theory. In the first step of Shoemaker's strategy, by removing the previous awareness condition, the identity criterion is entirely abandoned. Then, by reinstating a weakened version of this condition, he obtains an apparently ideal concept of memory: our common “strong remembering” is understood as a form of “weak remembering” characterized by a weakened causal chain: one not strong

enough to presuppose the identity of the rememberer but sufficient to preserve a minimal version of the “special access” individuals have to their personal past histories.

According to Shoemaker’s (2003a) account of genuine remembering in our world, it is clear that he regards the first person not only as a “singular referring expression” (Shoemaker, 2003b: 6) but also considers memory-based ‘I’-thoughts to be self-referential. He emphasizes the importance of causality in memory theories, which explains why he rejected the stronger version of quasi-memory. However, as Shoemaker (2003a: 46) himself points out, in a world where “strong remembering” can logically differ from “weak remembering”, and where one can weakly remember an action from the inside, from the first-person perspective, that memory will not be logically IEM but only de facto IEM, meaning IEM holds only “in normal, real-world conditions but could fail in hypothetical or abnormal scenarios (e.g., hallucinations or extreme memory distortions)” (Coliva & Palmira, 2025: 3).

Shoemaker’s move in conceiving the concept of causal quasi-remembering is not only incompatible with a strong logical IEM – “that holds across all possible worlds” (Coliva & Palmira, 2025: 3) – but also cannot be cleanly separated from the causal condition that relies on memory-based ‘I’-thoughts, the identity criterion. This second problem is further illustrated by Parfit’s (1984) definition of quasi-memory, in which he replaces the seemingly stricter ‘psychological continuity condition’ of Lockean personal identity with the looser concept of ‘psychological connectedness condition’. This shift enlarges the causal condition from requiring a stream of continuous consciousness to one that accepts temporary pauses and connections between discrete pieces of consciousness.

These moves do not appear very attractive for an attempt to completely eliminate the liability of circularity. Although memory causation is not solely about the identity of the first person, there is a strong reliance on identity in its concept. Versions of quasi-memory with a slightly loosened causal condition thus seem less compelling than the causality-free strong version of quasi-memory, which aligns with the same level of commitment to identity-lessness as Anscombe’s view on the first person.

In fact, if Anscombe is correct about the reference-lessness of the first person, then no causal version of quasi-memory can be sustained. As long as I use ‘I’ in quasi-memory claims, the subjective character of ‘I’-use grants me a (sometimes false) epistemic authority to take the apparent (quasi-)memory as my own personal past history. This observation aligns with

4 The First Person, IEM, and Quasi-Memory

Shoemaker's conclusion – although reached via the causal version of quasi-memory – that the first person is not logically but merely *de facto* IEM.

Furthermore, the question of IEM seems to place the present discussion in a dilemma: either preserving logical immunity requires maintaining first-person identity in memory-based 'I'-thought, thus risking a return to circularity, or it allows the removal of first-person identity from memory-based 'I'-thought, but then we face similar difficulties as with the stronger version of quasi-remembering regarding the epistemic capacity to provide positive information about the world.

Now that I have argued that the slight loosening of the causal condition isn't sufficient for Lockean personal identity to be safely explained by a causally conceptualized memory, and that such a causal concept of quasi-memory isn't compatible with Anscombe's reference-less 'I'-use, let us turn more closely to the philosophical significance of IEM. In the next section, I will demonstrate different kinds of IEM and defend a view inspired by Anscombe's claim that the special (*de facto*) IEM in memory-based 'I'-thought arises from the nature of the first person's 'capacity to say' within the linguistic structure, rather than from the use of 'I' actually identifying a past entity identical to the entity presently using 'I' either in speech or in thought to recall a past event or experience. In a way, we might say that the use of 'I' does not, and need not, identify with an entity because functionally it originates from the subject who exercises the function of 'I'.¹⁶

4.3 IEM and Personal Memory

IEM is the special power by which the objects of a proposition are protected from EM, and this immunity comes in different forms because EM itself takes various forms. A thinking subject can be mistaken about the identity of various things when forming beliefs, which, when properly justified, become knowledge. Depending on which objects of a proposition could potentially be misidentified, philosophers categorize different kinds of EM. In earlier sections, we have already mentioned two basic kinds of IEM: *de facto* and logical. The first is

¹⁶ This functionalist idea is heavily influenced by Fernández's (2019) functionalist self-referential memory theory; however, the view of 'I'-use that I support does not include a causal origin grounded in the identity of the object referred to by 'I', but instead includes the subject itself as the epistemic justification for the 'I'-function – the special 'capacity to say' that accompanies the use of 'I' without needing to explain the identity of any particular object.

effective in our actual world, whereas the second holds across all logically possible worlds. But that is not all.

Regarding the first person, according to Wittgenstein (1958), there are two kinds of 'I' uses: (1) the use as object, and (2) the use as subject. He argues that only the second kind – the 'I' use as subject – is IEM. Pryor (1999: 280) further elaborates that “all introspectively based beliefs about oneself” (corresponding to the “use as subject”) will be IEM, whereas many (though not all) “demonstrative beliefs and beliefs about one’s bodily state” (corresponding to the “use as object”) will be vulnerable to EMs. For example, I cannot be mistaken about the first person (as subject) when I say “I am in pain”, “I see a canary”, and “I hear Bach on the radio”. However, I can be mistaken about the first person (as object) when I say “My arm is broken”, “I’m wearing a blue hat”, and “My hair looks horrible”.

Pryor (1999) then elaborates a kind of IEM related to ‘de re’ misidentification, that is, when the misidentification concerns the identity of a particular person or object, rather than being about a description or proposition (de dicto). In this kind of de re misidentification, the description is correctly attributed to either the referring object or the referred object, but the identification between the referring and the referred objects fails.¹⁷ There is another kind of misidentification regarding the attribution of descriptions or properties of an object: while the properties are true of some particular referred object, the referring object is misidentified; in other words, the properties are attributed to the wrong particular object.

Pryor (1999: 281) names this second kind of error ‘which-object misidentification’, or ‘*wh*-misidentification’ for short. In de re misidentifications, I have two beliefs: one that a particular object *x* has the property *F*, and another that *y* has the same property *F*, and I take *y* to be the same object as *x*. I misidentify *y* as *x* when announcing “*y* is *F*”.¹⁸ The identity relation is misidentified. For example, if I see a man (*y*), think that he is my friend Sam (*x*), and tell the person next to me while pointing to the man, “There is Sam, the boy I grew up playing with (*F*)”, but it turns out the man isn’t Sam, this is a case of de re misidentification where I misidentify the man as my friend Sam.

¹⁷ By ‘referred object’ I mean the referent of the proposition, and by ‘referring object’ I mean the object that is considered to be, or thought to be, the referent in the proposition. In de re misidentifications, the referring object is misidentified as the referent (the actually referred object), but it is not.

¹⁸ In this case, *y* would be what I mean by ‘referring object’ and *x* the ‘referred (particular) object’.

4 The First Person, IEM, and Quasi-Memory

However, in *wh*-misidentification, I form only one belief that object *x* has the property *F*, but fail to identify which particular thing *x* is. For example, it smells funny (*F*) in a crowded café; I think it must be the dog resting in front of me who smells funny, and think to myself, “The dog smells funny”, but it turns out that it was in fact his owner (*x*) who smells funny. In this case, I’m *wh*-misidentifying the object that smells funny by attaching a description to the wrong object.

In *de re* misidentification, there is a prior epistemically justified belief about the (referred) object with the correctly attributed description, but I misidentify it as the (referring) object. In *wh*-misidentification, no prior epistemic belief is either formed or justified; the misidentification happens during the formation of the belief that *x* is *F*, whereas *F* is wrongly attributed to *x*, the (referred) object.

It seems that, regarding *de re* misidentification, the ‘I’ use as subject is immune to *wh*-misidentification, although this is not the case for the ‘I’ use as object. Examples such as “I am in pain”, “I see a canary”, and “I hear Bach on the radio” are beliefs formed on the basis of introspection, and there is no need for other grounds to justify their formation. Moreover, there is no defeater that can offer contrary evidence against these introspective beliefs.¹⁹ This is highly compatible with our earlier observation that ‘I’-thoughts are subjectively characterized by an immediate, direct, and straightforward capacity to say, needing no justification other than the very function of ‘I’ itself. Pryor (1999: 286) correctly judges that *wh*-misidentification is epistemologically more central and interesting because it occurs during the formation of the belief about an object. This point can be emphasized by recognizing as immunity to *wh*-misidentification both Wittgenstein’s remark that asking “Are you sure it’s you who have pains?” would be nonsensical, and Shoemaker’s claim that “there is no room” for the thought “Someone is hungry all right, but is it me?” (Pryor, 1999: 287). Even though Shoemaker does not explicitly distinguish between *de re* and *wh*-misidentifications when introducing the notion of IEM, his (2003b) examples of IEM do seem to adopt the more basic epistemic status of immunity to *wh*-misidentification.

¹⁹ Pryor (1999: 284) distinguishes two kinds of defeaters – additive and undercutting – to clarify which kinds of propositions are vulnerable to *wh*-misidentification. Additive defeaters provide direct, positive evidence against the proposition in question, thereby defeating the belief. In contrast, undercutting defeaters challenge the basis or justification of the belief by offering an alternative explanation for the grounds on which the belief was formed. Vulnerability to *wh*-misidentification, according to Pryor, is vulnerability to undercutting defeaters.

The reason Shoemaker chooses the second, weaker version of quasi-memory – one with a causal condition similar to that of genuine memory in our world – is that the stronger version of quasi-memory is vulnerable to *wh*-misidentification in first-person beliefs. The elaboration of quasi-remembering shows that “the first-person beliefs about the past that we base on our apparent memories are not really immune to error through misidentification”. However, since in our world all of our memories belong to our own past, “our memory-based beliefs enjoy a de facto immunity to *wh*-misidentification” (Pryor, 1999: 289). Evans (1982) argues against Shoemaker that even in quasi-remembering, the use of the first person is not merely de facto IEM, but logically IEM, due to the ‘identity-less-ness’ of ‘I’.

Even though the ‘identity-less’ view proposed by Evans has a particularly compelling appeal, it does not seem to withstand the scrutiny demonstrated by Pryor (1999: 290–291). Taking Shoemaker’s side in the Shoemaker–Evans debate, Pryor argues that our first-person memory-based beliefs do not possess logical IEM. To see why human memory is not logically IEM, it suffices to consider the ‘I’-thought in cases of quasi-memory, while keeping in mind the basic epistemic power of ‘I’: namely, the subject’s ‘capacity to say’ that is intrinsic to the function of ‘I’-use. In possible worlds where quasi-memories can concern someone else’s past experiences, if I say “I went to Rome last week”, I may have an apparent memory of being in Rome last week from a first-person perspective and from the inside. Yet there is no way of knowing whether that quasi-memory belongs to someone else, since the use of ‘I’ casts me as the subject of that memory and confers upon me the (in this case, false) authority to claim it as mine. Therefore, I am vulnerable to *wh*-misidentification as long as I lack either positive information confirming the correct identification or a defeater proving that the quasi-memory is false evidence for a belief about my personal past. Since there is no guaranteed access to either kind of information, we can conclude that remembering – understood as a form of quasi-remembering – is, in fact, only de facto and not logically IEM.

I have argued that the use of ‘I’ is, in itself, functionally epistemic for the subject using it, in that it requires no external justification. However, this view is not without challenges, the most obvious being this: if ‘I’ does not refer to an identity or a particular object, how can my memory-based ‘I’-thoughts grant me any special access to my own personal past? Shoemaker (2003a) suggests that in worlds where quasi-memories exist, identity may not be as significant as it is in our world. If we were to experience fusions and fissions on a daily or even hourly basis, we might well lose track of our identity by the end of the week. I believe

4 The First Person, IEM, and Quasi-Memory

this is what inspired Parfit's (1984) survival theory of personal identity, according to which identity is not what ultimately matters for survival. In light of these possible worlds, personal identity may require an entirely new theoretical framework: one in which psychological continuity, and therefore memory, still plays a fundamental role, but might not be sufficient. If adopting such a view becomes necessary as a compromise for achieving a non-causal and object-less theory of memory, it may imply the eventual relevance of the bodily criterion. As we have seen with Anscombe, the strength of her view of 'I' without reference to an object lies in its deep connection with non-observational knowledge, which depends heavily on bodily movement and sensation. This would mark a significant departure from neo-Lockean theories, and a move toward hybrid views.

In that case, a conclusion can be drawn: a pure neo-Lockean theory of personal identity seems to depend on causal theories of memory, as it requires the self-referential character of memory to account for the sameness of persons. However, such a theory – one that claims memory to be not only necessary but also sufficient for personal identity – is vulnerable to the charge of circularity. If it adopts a non-causal theory, such as simulationism, the challenge becomes whether psychological continuity can still be strong enough to ground identity without already presupposing it. If it maintains a causal condition, it risks circular reasoning by grounding identity in a memory that itself requires identity. This dilemma mirrors the one posed by the problem of IEM discussed at the end of the previous section.

At this stage, part of my working hypothesis can be confirmed: the neo-Lockean theory remains conceptually entangled with the causal theory of memory, and relaxing the causal requirement does not automatically solve the circularity problem. That said, some versions of simulationism may attempt to embed personal identity into the functional structure of memory without relying on a causal link, and this theoretical space should not be ruled out prematurely. For this reason, a viable theory of personal identity may still be informed by neo-Lockean insights, but must go beyond the classical memory-based framework to avoid circularity while preserving the role of psychological continuity.

This conclusion is reached by adopting the strategy of endorsing an object-less use of 'I' in memory claims: specifically, in possible worlds where quasi-remembering occurs, and where memory-based 'I'-thoughts are not immune to IEM. If, logically speaking – that is, across all possible worlds – personal identity proves to be epistemically irrelevant to the

functioning of memory, then it may be time to move beyond treating identity as a necessary criterion in theories of memory.

4.4 Conclusion

In this chapter, I have examined Anscombe's view on the first person, drawing from her notions of non-observational knowledge and practical reasoning. I argued that her object-less conception of 'I' offers valuable insight into how we should approach Shoemaker's two versions of quasi-memory. By unpacking the notion of IEM, I aligned with Shoemaker and Pryor in concluding that memory-based 'I'-thoughts enjoy only a *de facto*, not a logical, immunity to misidentification.

Of course, the identity-less view of memory raises further challenges. For instance, if there is no metaphysical self as a persisting thinking subject, how is self-consciousness preserved? Does selfhood become a mere bundle of past experiences? Such a view seems unattractive, as it implies a passive accumulation of experiences without agency. Could it instead be a bundle of metacognitive functions? If so, these functions would require a physical basis in the brain, leading us step by step toward the core intuition behind the simulation theory of memory.

Ultimately, I conclude that a purely neo-Lockean theory of personal identity is unlikely to suffice. It may be more fruitful to move beyond this view and acknowledge the necessity of certain bodily criteria. To support this shift, I will first address an additional phenomenological argument for the identity-less character of the first person in memory before exploring how the simulation theory of memory aligns with this perspective. Finally, I will consider the compatibility of a neo-Lockean theory of personal identity with a post-causalist account of memory.

Part III: Toward a New Framework

This final part of the thesis aims to develop a new approach to personal identity grounded in the phenomenology of memory and the simulationist theory of memory. Moving beyond traditional neo-Lockean and causal theories, I focus on the interplay between episodic remembering, imagining, and the distinctive feeling of mineness that characterizes our first-person experience of memory. By exploring how episodic thinking unfolds within what I term the “episodic tube” – a conceptual space that delimits genuine episodic mental states – I investigate how our mental representations shape and sustain our sense of self over time. This framework reconciles the dynamic and reconstructive nature of memory with the continuity necessary for personal identity, emphasizing the importance of psychological coherence rather than strict causal links.

The two chapters that follow develop this new framework in detail. The first examines how the sense of ownership over memories arises independently from factual accuracy or causal connection, challenging assumptions about what constitutes genuine episodic remembering. The second builds on this by proposing a hybrid, neo-Lockean account of personal identity that relies on simulationist memory and the metacognitive systems that produce it. This account circumvents traditional circularity problems and accommodates complex phenomena such as fusion, fission, and quasi-memory, thereby offering a more nuanced and flexible metaphysical theory of the self-grounded in lived experience.

5 The Feeling of Mineness and Remembering

The main objective of this thesis is to address the following question: Can the circularity problem be avoided by adopting the simulation theory of memory? As I argued in Chapter 3, the circularity in neo-Lockean theories arises because causalist memory presupposes personal identity. Thus, it is problematic to claim that memory alone maintains the continuity of a person that constitutes their numerical identity over time. Although the causal theory of memory has been highly influential in the philosophy of memory, it struggles to adequately account for personal identity. The difficulty lies in its causal condition. In response, some philosophers – including notably Shoemaker and Parfit – have adopted strategies that loosen this strong causal connectedness, most significantly through the concept of quasi-memory.

We have seen the conceptual value of quasi-remembering and quasi-memory in the preceding chapter. However, regarding personal identity, Schechtman (1990) strengthens the circularity objection by denying the logical conceivability of genuine quasi-memory, as opposed to mere quasi-memory. She creates a dilemma by claiming that a quasi-memory candidate will necessarily be either delusional or presuppose personal identity. The fundamental reason for this, Schechtman argues, is that the sense of ownership – one specific aspect of the phenomenology of remembering – is inseparable from the memory content.

Klein and Nichols (2012) respond to Schechtman's dilemma by showing that the sense of mineness – the feeling of ownership – can, in fact, be detached from genuine memory, using the well-known case of patient R.B. as an illustration. The fact that R.B. was able to articulate his own past experiences and later “own back” some of his previously disowned post-accident memories provides evidence that the occurrent mental images he reported should be considered genuine episodic memories rather than mere imaginings. In this case, it follows that genuinely remembering a past experience episodically does not require the sense of mineness; the feeling of ownership is only a contingent feature of memory (Klein & Nichols, 2012: 689).

Fernández (2019) agrees that the reports of patient R.B. should be taken seriously for their theoretical value, since the structure of the phrase “I have such and such a mental state,

but it is not mine” is not unique to R.B. or to memory. It has also been observed in other psychological and neurological cases. Nevertheless, Fernández argues that Klein and Nichols’s (2012) ‘identification model’ does not faithfully capture R.B.’s verbal reports. Through a linguistic analysis of R.B.’s statements, Fernández concludes that R.B. is not experiencing an identity issue. Rather, the feeling of ownership is not about resembling a previous person who experienced the remembered events, but about a sense of alignment between the current mental images and the veridicality of the past event. He refers to this interpretation as the ‘endorsement model’. Fernández’s endorsement model presents a challenge to the project of developing a theory of memory that avoids presupposing personal identity. By treating R.B.’s disowned memories as endorsed and therefore as genuinely his, Fernández effectively reintroduces the notion of identity into the analysis of memory: a move that runs counter to Klein and Nichols’s attempt to separate memory from identity conditions.

On the other hand, with regard to the epistemology of memory, Fernández (2019) also challenges the notion of quasi-memory as a threat to memory’s characteristic of IEM. In discussing memory judgments, Fernández objects to the challenge that Shoemaker (2003a) poses to memory’s IEM via quasi-memory by articulating an argument similar to Schechtman’s. He points out that whether memory enjoys IEM hinges precisely on whether memory is silent regarding the feeling of mineness. From there, he simply reaffirms his earlier analytic conclusion: namely, that memory is not silent on this feeling. Since the previous argument (in support of quasi-memory) was refuted, this one, which depends on it, is likewise rejected for the same reasons.

While Fernández’s account treats memory as self-referential – that is, a memory represents an experience as having happened to the rememberer – it does not require a causal connection in the strong metaphysical sense used by traditional causal theories of memory. Instead, the causal link in his view serves an epistemic role: it supports the subject’s endorsement of the memory, without presupposing identity between the subject of the original experience and the rememberer. This distinction is crucial for theories like quasi-memory, which aim to retain the structural features of memory without relying on personal identity. To fully accommodate such possibilities, we may need a theory of memory that preserves the reflexive character of memory content while decoupling it from any commitment to diachronic subject-identity.

In this chapter, I will argue against Schechtman and in favor of Klein and Nichols that the feeling of mineness is only a contingent feature of remembering. This is what makes genuine quasi-memory conceivable, especially when combined with the argument derived from Anscombe's first-person thesis. I will partially endorse Fernández's endorsement model of mineness but reject his claim that the feeling of ownership is a necessary condition for genuine episodic remembering, drawing on empirical evidence from cases of non-believed memories. However, if the feeling of ownership is indeed contingent, this raises the risk of misidentifying remembering with imagining: a problem that will open the discussion for the next chapter.

5.1 Schechtman's dilemma

In response to the concept of quasi-memory, Schechtman (1990) chose Parfit's most radical version as her object of analysis and reinforced the circularity objection by arguing that any quasi-memory candidate will either be delusional or presuppose personal identity. As such, it still falls into the inherent circularity of Locke's original theory of personal identity, in which memory plays a fundamental role in constituting consciousness. She argues that "psychological continuity does not and cannot provide the sort of criterion of personal identity which identity theorists wish to provide" (Schechtman, 1990: 71). Her main supporting argument is that memory is not a mental state that can be isolated from other psychological features – most notably, the distinctive feeling that an episodic memory is one's own, a sense of ownership she claims cannot be detached from the memory content. To support her claim, Schechtman draws on a memory example provided by Edward Casey in *Remembering: A Phenomenological Study* (2000: 25–26; Schechtman, 1990: 80). In Casey's detailed description of a sequence of memory episodes, the reader is drawn into his first-person experience of a casual family outing. We are placed in his shoes, reliving moments spent with his family, with whom he shares a special and intimate bond.

To consider Schechtman's claim more illustratively – that Casey's sequence of memory episodes cannot be quasi-remembered correctly and non-delusionally – let us borrow Jane's body for a thought experiment. There are two possible scenarios: either Jane receives only the memory content from Casey's sequence of episodic memories, or she receives every psychological feature relevant to that sequence (Schechtman, 1990: 82). In the first case, the memory content would appear drastically out of context. Consider the following variation on

5 The Feeling of Mineness and Remembering

the ‘Venetian memories’ case: Jane receives Paul’s memory traces containing memorial representations, one of which is a distinct episode involving a conversation with a local who told a joke in Italian. While Paul is a fluent Italian speaker and found the joke amusing at the time, how is Jane – a non-Italian speaker – supposed to perceive this apparent memory? Is she to recall the joke-teller’s speech as incomprehensible gibberish? After all, the Italian phrases, the speaker’s expressions, and the surrounding context arguably constitute the memory content. We must agree with Schechtman that this appears nonsensical. Such an apparent memory could hardly qualify as memory at all: it would be better described as a collection of alien and meaningless mental images with no intelligible content for Jane.

By analogy, if Jane were to receive only the memory content of Casey’s sequence of episodic memories, that content alone would similarly make no sense. Quasi-remembering glancing at Casey’s children during a film would seem uninteresting to Jane: not only because she doesn’t know Casey’s children, but, more fundamentally, because she doesn’t even know who Casey is. This inherited memory sequence, stripped of its relevant psychological context, would not yield a meaningful recollection and should therefore be considered a delusional episode rather than any genuine form of memory.

I propose an objection to the assumption that remembering the content of an Italian joke requires the (quasi-)rememberer to understand Italian. It is not difficult to imagine a person whose first language was Mandarin, acquired in early childhood. At the age of eight, this person moves to an English-speaking country and, in a dramatic turn of life, never again encounters the need to use Mandarin. As a result, they gradually lose their ability to speak the language by the age of twenty-five, despite awkward attempts to pronounce a few words when teased for being a “Mandarin-looking” person who does not know the language.

Does this mean the person would be unable to remember memory content from early childhood that involved Mandarin? This seems counterintuitive. Suppose there was a particularly joyful joke told by a kind adult that amused me as a child: at a time when I still spoke the language I no longer speak today. It seems implausible to say that I would be unable to remember the feeling of amusement associated with that event. In fact, it seems more likely than not that I might be able to remember the meaning of the joke, even if I cannot recall the exact phrasing in Mandarin, or even any specific Mandarin words.

This experience is likely familiar to those who underwent a permanent international relocation during childhood, becoming acquainted with certain languages only to lose that

acquaintance at a later stage. It seems unnatural to claim that not remembering the precise linguistic form of a memory undermines one's ability to remember the meaning of a past event. Therefore, it appears unnecessary to require understanding the language used in a past event in order for a subject to genuinely remember it episodically.

However, this argument might be challenged on the grounds that, unlike the ex-Mandarin speaker, Jane never knew Italian and therefore lacks any cognitive familiarity with the joke in Paul's apparent memory. I believe this merely reinforces Schechtman's (1990) point about the implausibility of memory when stripped of other psychological features. This view is also compatible with Anscombe's (1975) account of first-person memory reports, discussed in the previous chapter: the use of 'I' in memory statements does not reflect an identification based on observation or inference, but rather a form of non-observational, practical knowledge. This knowledge arises through the psychological features that accompany the act of remembering itself: such as affective tone, imagery, or a sense of agency. Without these accompanying features, there is no immediate basis for the subject to recognize themselves as the one who originally experienced the remembered event.

This leads us to Schechtman's second possible scenario, in which Jane inherits not only the memory content of Casey's sequence of episodic memories but also every psychological feature attached to it: a situation in which "Casey's memory reproduces in Jane exactly, with all of its personal elements" (Schechtman, 1990: 83). By "personal elements", Schechtman means any of Casey's personal associations with this episodic memory, including his sentiments toward his wife and children, his preferences regarding the film, and his mood during the experience. Schechtman finds this scenario unlikely because, in such a case, Jane would effectively take Casey's wife and children to be her own, since "the mineness of the experience seems to be part of the content of the memory" (Schechtman, 1990: 83). She emphasizes this point by arguing that a "memory that is familiar and nostalgic to Casey [is in this case] foreign and fantastic to Jane" (Schechtman, 1990: 83, emphasis added).

However, it seems that what Schechtman is really suggesting is that the main reason why mineness appears inseparable from Casey's memory is that, without the feeling of ownership, the quasi-rememberer would be unable to make sense of the psychological features accompanying the memory content. Jane would not be able to understand the feelings generated by Casey's psyche unless she took Casey to be herself. This argument aims to underscore the attachment between the feeling of mineness and memory. But this raises

5 The Feeling of Mineness and Remembering

an interesting question: on what basis are we assuming that making sense of the full set of psychological features attached to a memory requires the feeling of mineness? In other words, why presume that a quasi-rememberer must take the quasi-memory to be their own in order to comprehend the meaning of a memory episode?

We have all certainly experienced the feeling of fully understanding other people's emotions in various circumstances: when the experience is highly relatable, when the story is vividly told, or when the narrative includes elements that resonate deeply within us. This happens often in our daily lives: for example, when we read a book whose main character undergoes an experience strikingly similar to our own, making the passage feel like a second reliving of our past. Or when we watch a 4K film in the cinema, where not only is the image quality so vivid it feels like we can almost touch the sets, but the storyline is so well crafted that we can easily create mental images as if we were the main character on those fantastic adventures. Granted, in these cases, we usually fully understand the psychological features accompanying the mental images – whether through written description or visual effects – and yet we rarely mistake ourselves for the originator of those mental images, no matter how compelling the story or how high the image quality. Thus, we can understand memory-like experiences without necessarily feeling that they are our own.

It seems, therefore, plausible to say that the feeling of mineness is not necessary for *understanding what happens* in a memory – that is, for grasping the structure and content of the episode – especially in the second scenario described by Schechtman, where all the psychological features associated with the memory are transferred to the quasi-rememberer. In such a case, Jane can quasi-remember Casey's evening family outing by comprehending the memory and making sense of the experience, even if she lacks the feeling of identity with Casey. However, what she may lack is the ability to situate that memory as *part of her own life*. Still, in the absence of third-party verification, the presence of all the psychological features from a first-person perspective may lead Jane to take the memory as her own. In this context, genuine quasi-memory need not be seen as delusional, as Schechtman suggests, but as a coherent phenomenological possibility: especially when we recognize, as argued in the previous chapter, that IEM is only de facto rather than a logically necessary feature of personal identity across all possible worlds.

From another perspective, we can say that Schechtman is challenging people's general capacity for compassion and empathy. Quasi-remembering doesn't require the quasi-

rememberer to feel identical to the person who experienced the event that resulted in an episodic memory; instead, it requires the quasi-rememberer to stand “in the (first-person) shoes of” the experiencer without necessarily having to recognize the feelings they now feel as ones obtained through their own sensory modalities. It isn’t unreasonable to find this difficult to conceptualize because such a plausible situation is counterintuitive and destabilizing. Having every psychological feature attached to the memory content transmitted to one’s mind – except for the one feature of mineness – seems like a very confusing situation at face value. Losing identity in memory is like losing our own benchmark of reality. If memory doesn’t necessarily contain identity, how do we know which episode we quasi-remember actually happened to us? The quasi-rememberers will then frequently question themselves about the source of these memory episodes. But despite its unsettling character, is this situation implausible?

Let us now turn to the case of patient R.B., whose experience provides a clinical illustration of quasi-memory-like phenomena. Though R.B. retains some memories of his personal past, these memories are often fragmented and lack the full first-person ownership typical of episodic memory, reflecting a state where memory content is preserved without the usual causal and subjective features. This case relates to quasi-memory not because R.B.’s unowned memories do not actually belong to him (they in fact are genuinely his), but rather because it demonstrates that one can possess memories without the accompanying first-personal feeling of remembering. In other words, R.B.’s case illustrates how memory content can persist even when the subjective experience or ‘mineness’ typical of episodic memory is diminished or absent.

5.2 R.B.’s Mineness-Less Memories

Klein and Nichols (2012) published a paper presenting a unique, unprecedented clinical case of a patient pseudonymized as “R.B.”, who claims to have lost the feeling of ownership for certain genuine memories from before a traumatic accident. They argue, based on R.B.’s verbal reports, that these memories are episodic rather than semantic, as they involve features characteristic of episodic remembering, such as the ability to be auto-cued (Donald, 1991; Klein & Nichols, 2012: 687) and mental time travel (Klein et al., 1996; Klein & Nichols, 2012: 688; Michaelian, 2016). For example, R.B. reported, “I can picture the scene perfectly clearly... [...] It did not feel like it was something that really had been a part of my life”, and

5 The Feeling of Mineness and Remembering

that “I did not ‘own’ any memories that came before my injury” (Klein & Nichols, 2012: 685–686). From this, Klein and Nichols emphasize that the phenomenology of remembering – that is, the feeling of mineness – is only a contingent feature of genuine episodic remembering.²⁰

Patient R.B.’s case has not been uncontroversial (Roache, 2016; Trakas, 2021). Roache (2016) especially pointed out the risks of overgeneralizing from R.B.’s case to broader theories of identity. However, as Fernández (2019) emphasized, the formulation of R.B.’s verbal report about his apparent memories before the accident is compatible with the structure of “I have such-and-such mental state, but it is not mine”. This structure, being compatible with clinical cases in psychiatry and neurology, such as delusional disorders, anarchic hand syndrome, and utilization behavior (Fernández, 2019: 128–132), should be enough reason to take such a formulation of memory claims seriously.

It seems possible to illustrate an analogy of quasi-remembering through R.B.’s experience. If quasi-memory does not necessarily require the punctual feeling of mineness, then it becomes even more plausible to imagine quasi-rememberers entertaining first-personal memory representations without experiencing themselves as the subject of the remembered event. R.B., for example, can recognize that the memories are *about him*, but does not experience them with a sense of first-personal ownership. Roache (2016) offers a challenge similar to Schechtman’s: that it is inconceivable for R.B. to retain all aspects of memory while lacking only the feeling of mineness: if such a dissociation were possible, she argues, it should be accompanied by other impairments in the memory system. However, as I argued in the previous section, genuine remembering involves understanding the memory content and taking it as one’s own through a first-personal mode of presentation – not necessarily through a felt sense of ownership. In this respect, the phenomenological feeling of owning the episode is not constitutive of genuine remembering, though it may often accompany it. Following Schechtman and Roache, one might still argue that, even if the feeling of mineness is not necessary, it is nonetheless a *de facto* feature of most genuine remembering in our world. In that case, should it not count as a *de facto* meaningful criterion of genuine remembering?

²⁰ In the phenomenology of memory literature, there are flourishing discussions concerning the metaphysics of the feeling of remembering, including aspects such as the ‘feeling of pastness’ (or ‘experience of time’) and the ‘epistemic feeling’ (Fernández, 2019; Perrin et al., 2020; Perrin & Sant’Anna, 2022). In this thesis, I focus primarily on the feeling of mineness, as it appears to be the most pertinent to personal identity in memory.

Fernández (2019) proposes an 'endorsement model' to interpret patient R.B.'s mineness-less memory reports, aiming to preserve the feeling of mineness as a necessary feature of genuine remembering. In this view, R.B.'s unowned memories are disqualified as genuine episodic memories.

5.3 Fernández on the Feeling of Ownership

Before proposing his model, Fernández (2019) first argues *via negativa* against Klein and Nichols's (2012) 'identification model' of interpreting R.B.'s reports. This model of mineness can be summarized as the view that "the experience of a memory of some scene as being the subject's own is the feeling of being identical with a past person; a person involved in the remembered scene" (Fernández, 2019: 113). The identification model can be further divided into two versions: the 'experiencer' and the 'object' versions. According to the experiencer version, the feeling of mineness arises from identifying the rememberer with the subject who perceptually experienced the remembered event. Conversely, the object version holds that the feeling of mineness comes from identifying the rememberer with the subject who was present at the past (now remembered) event. The difference between these two versions can be subtle, but Fernández's critique helps clarify the distinction.

Here are some identified difficulties with the two versions of the identification model. Fernández draws our attention to the pronouns used by R.B. in his disowned memory claims. For example, in report 1, he states "I was remembering scenes"; in report 2, "having been"; in report 3, "studying with my friends"; and in report 7, "myself walking into the room" (Fernández, 2019: 114–116, *emphases added*). The issue raised by Fernández is the following: R.B. does not appear to have any difficulty identifying with the object of his disowned memories, since he refers to that object using the first person. This suggests that the identification model, in its object version, fails to account for the experience of disowning memories described by R.B. In the experiencer version, Fernández highlights the verbs used by R.B., such as "relive" and "re-running", which appear in reports 3 and 4. In report 4, R.B. also states that he had "a sense of being there, at MIT, in the lounge" (Fernández, 2019: 115). According to Fernández, these expressions are problematic: they indicate that R.B. does not experience any difficulty identifying with the person who experienced the remembered scenes. The repeated use of verbs with the prefix "re-" suggests, in fact, that R.B. perceives himself as having genuinely lived through those experiences in the past.

5 The Feeling of Mineness and Remembering

Via affirmativa, the 'endorsement model' proposed by Fernández is designed to overcome the two main limitations of the identification model. According to Fernández (2019: 119–123), the feeling of ownership is not merely an identification with an object or an experiencer from the past, but rather the feeling that a remembered episode reasonably matches the past. More precisely, he claims that the feeling that a memory instance belongs to a subject is equivalent to the feeling that the remembered episode faithfully corresponds to a past event. To better understand this model, it is necessary to return to Fernández's reflexive thesis on the content of memory, according to which, under normal conditions, a memory instance represents a past perceptual experience that satisfies three essential criteria: (1) it caused the act of remembering, (2) it was truthful at the time the experience occurred, and (3) it is intentionally represented in the content of the memory episode.

While this model does not explicitly invoke personal identity as a requirement, one might argue that the first criterion implicitly presupposes identity: for a past experience to cause the present mnemonic state, it must have been the subject's own past experience. However, Fernández distinguishes between two levels in memory: the metaphysical level, which concerns what makes a mental state a memory at all, and the intentional level, which concerns what memory represents. Causality plays a different role at each level: functionally grounding memory at the metaphysical level, while being reflexively represented within the memory's content at the intentional level. In this light, while the functionalist causal condition is necessary at the metaphysical level to distinguish genuine memories from other mental states, it plays no role in the phenomenology of memory itself. On Fernández's account, the content of memory represents a past experience reflexively, without including any felt sense of causality or identity with the past self. As a result, the endorsement model may not entirely escape metaphysical commitment to the identity criterion, even as it reframes the phenomenology of memory in terms of content endorsement rather than identity-based recognition.

According to my conclusion from the last chapter, which follows from Anscombe's view on the first person, there is no need to identify an object, as Fernández also suggests, when using 'I'-claims in memory reports. It doesn't make sense for the 'I'-user to try to identify with some object that experienced the past event. Fernández's functionalism, built on the reflexive thesis as a non-causal and therefore identity-less memory theory, supports an endorsement model of the feeling of ownership that provides a reading of R.B.'s reports without implying

personal identity in remembering. The feeling of mineness is not the feeling of being the same person who experienced the event, but rather the feeling that my current apparent memory truthfully reports positive information about the past.

Through this reading, Fernández's endorsement model offers an interpretation of R.B.'s memory reports as genuine memories that lack the typical feeling of ownership: "if the endorsement model is right, then R.B. believes that he is having certain memories, but he does not have the feeling that the scenes represented in those memories happened in the past. [...] then he should not have the feeling that he is remembering those scenes" (2019: 121). In this way, Fernández explains how R.B.'s memories can be intact in content yet devoid of the phenomenological identification model of "feeling of mineness". Rather than disqualifying these "unowned" memories as non-genuine, the endorsement model accounts for their altered phenomenology. Nevertheless, while the model captures the feeling of mineness as an important aspect of our experience of remembering, this phenomenological feature should not be taken as a necessary criterion for genuine memory.

Trakas (2021) proposes reading Fernández as endorsing the view that, for an apparent memory – that is, a current mental representation – to feel like it matches the past, the content of that memory must include its own causal origin, namely the past perceptual experience. According to the reflexive view, only instances that represent their own causal origin qualify as genuine episodic memories. Trakas argues that, since it is possible to feel mineness and have the impression that our apparent memory matches the past even when it does not, the endorsement model is problematic. However, this issue arises not from the endorsement model itself, but from the tension between the reflexive content condition and the phenomenology of endorsement. While Fernández maintains that genuine memory content includes a causal and truthful origin, the endorsement model characterizes the feeling of ownership or "mineness" as a phenomenological experience, which can occur even in false memories. In other words, although the ideal content of a memory reflexively includes its causal origin, the feeling that a memory matches the past can sometimes be experienced in memories that are not veridical.

Trakas (2021) offers symmetrical examples to argue that the endorsement model cannot fully capture the feeling of ownership. Not only can we endorse false memories, but we can also fail to endorse veridical memories. On one side of this symmetry, many of us have personal examples like those Trakas describes: believing that we genuinely and vividly

5 The Feeling of Mineness and Remembering

remember episodes from our childhood, even though our parents have told us repeatedly that these events did not actually happen. Trakas claims that she stopped endorsing one such memory once she came to understand that the episode in question did not occur. However, we can also easily find examples where people continue to endorse such memories despite being told by their parents that these memories are false. On the other side of the symmetry, empirical evidence from studies on non-believed memories shows that veridical memories can exist without endorsement: i.e., as apparent memories lacking the sense of mineness (Trakas, 2021: 161–162; see also Mazzoni et al., 2010; Otgaar et al., 2014; Otgaar et al., 2019).

Table 2 Endorsement as the feeling of mineness in episodic memory.

	Endorsed	Unendorsed
Veridical Memory	Believed Veridical Memory	Non-Believed Veridical Memory
False Memory	Believed False Memory	Non-Believed False Memory

It seems, then, that the problem is not with the endorsement model itself, as Trakas (2021) claims, but rather with the compatibility between the endorsement model and the reflexive view in their attempt to account for the phenomena of remembering. Faced with the symmetrical examples of non-believed veridical memories and believed false memories, Fernández needs to explain how the reflexive view can account for apparent memories that do not contain any past perceptual experience as their causal origin, yet still allow the subject to endorse them and feel ownership. Is it because the memory content is similar enough to a forgotten past perceptual experience? Or does past perceptual experience also include past mental perceptual experiences that incorporate certain forms of imagination? These questions remain unanswered, and as Trakas (2021) points out, Fernández might require more empirical evidence to support his theoretical framework.

Meanwhile, it seems plausible to establish that the general feeling of mineness, which appears to accompany our endorsement of certain mental representations, is only a contingent feature of episodic remembering, as claimed by Klein and Nichols (2012).

5.4 Conclusion

In this chapter, I argued against Schechtman by claiming that the feeling of mineness is only a contingent feature of remembering, and that genuine quasi-memory is conceivable. I discussed Fernández's view on the metaphysics of the feeling of mineness, which he interprets as a form of endorsement rather than identification, as proposed by Klein and Nichols (2012).

Identity-Less Personal Past

While I have argued in favor of the endorsement model, I also raised concerns about its compatibility with Fernández's reflexive thesis of memory content, which struggles to account for certain cases of endorsement in memory. Ultimately, I established the plausibility of endorsement – as the feeling of mineness – being a contingent feature of genuine remembering.

If the feeling of ownership over memory episodes is as unstable as we have seen – sometimes present in false memories and absent in genuine ones – then what kind of theoretical framework can reconcile this phenomenological inconsistency with the continued epistemic reliability of memory as a basis for knowledge and personal identity? In the next chapter, I will explore this question further by applying Fernández's test of theoretical compatibility to a range of cases, aiming to assess which theory of memory best supports a neo-Lockean account of personal identity.

6 Identity-Less Memory-Based Personal Identity

Philosophers have long grappled with defining what remembering and memory truly are, engaging deeply with both philosophical concepts and empirical studies. The difficulty of this research lies in the fact that, despite recent advances in neuroscience – such as studies on neuronal firing patterns offering inductive insights into what remembering entails (see Sugar & Moser, 2019; Roux et al., 2022; Kolibius, 2023) – neither the act of remembering nor its mnemonic product is directly observable. This challenge highlights why the investigation of memory matters not only to philosophers but also to scientists, whose empirical contributions help bring conceptual ideas closer to observable facts through a scientific lens.

Scholars have been interested in memory for such a long time that questions of different kinds and around various aspects of memory have flourished. For example, the phenomenology of memory has become a popular subject in recent years, whereas discussions about memory traces have been ongoing for quite a while. The epistemology of memory typically emerged alongside a linguistic approach, while debates over memory representation arise from a more imagistic perspective. Regardless of the increasingly detailed subcategories in memory research, the nature – or metaphysics – of memory remains the foundational basis for various theoretical frameworks. In Fernández's (2019) recent work on a self-referential account of memory, he develops a functionalist view of memory's metaphysics by emphasizing a version of memory defined by its function, which he calls the 'mnemonic role' (2019: 49). However, there is room for objections and methodological critique regarding Fernández's promotion of functionalism. He relies on a purely intuitive method when qualifying memory instances, and, raising a pressing need for discussion, the epistemic relevance criterion he uses to distinguish functionalism from rival theories appears to invite certain doubts.

In this chapter, I will first analyze Fernández's (2019) test and results concerning the compatibility of different memory theories with our intuitions about which instances qualify as memory. I examine Fernández's step-by-step strategy of systematically dismissing each of his selected opponents in favor of his functionalist view. This process raises questions about

the relevance of the epistemic relevance criterion he uses to support functionalism's success. In the second section, I consider epistemic relevance as presented and detailed by Debus (2010), and argue in favor of the simulation theory of memory by challenging the relevance of epistemic relevance in defining genuine episodic remembering. Since episodic remembering is only part of a broader episodic construction system that includes episodic imagining, epistemic relevance does not seem as necessarily crucial as Debus and Fernández claim. However, even if it were necessary, simulationism appears to be as compatible with the epistemic relevance condition as other memory theories. In the third section, I illustrate a personal identity framework based on simulationist memory, highlighting its promise both for its identity-less nature and its epistemic strength. This will be followed by a brief conclusion.

6.1 A Compatibility Test for Memory Theories

Fernández presents a chapter on the functional character of memory in a collective work (Fernández, 2018), which later became a chapter in his own book on a self-referential account of memory (Fernández, 2019). In this chapter, he develops functionalism as a metaphysical theory of memory and designs a test for different memory theories to pass, before asserting the prominence of the functionalist theory of memory (FTM) over others. The competing accounts he considers are the causal theory of memory (CTM) and the narrative theory of memory (NTM). For the sake of completeness, I will add the simulation theory of memory (STM) to the list of rival theories.²¹ The criteria for passing Fernández's test concern these theories' compatibility with our intuitions in several key areas: content generation, isolation of episodes, confabulation, and epistemic irrelevance.

Content generation, in a broad sense, means any kind of additional information that was not part of the originally encoded representation of an event but is added to the retrieved memory after reconstruction. Fernández illustrates this with a case where he mentally represents his father wearing a belt with a silver buckle during a hunting session, which he knows he could not have perceived from his original experience due to the limited visual angle from where he was standing at the time (2018: 56). This is an example of radical content generation, where the added information was not even implicitly present in the original experience but comes from other sources of knowledge (in this case, Fernández knows

²¹ The nature, development, and respective potential of these memory theories in relation to a non-circular neo-Lockean account are discussed in greater detail in Chapter 3 of this thesis.

6 Identity-Less Memory-Based Personal Identity

semantically about the look of his father's belt and silver buckle, and therefore adds post hoc information to the reconstructed memory of the original event). The traditional version of CTM is obviously incompatible with radical content generation since it allows only the retrieval of information from the originally encoded representation of the past experience. The classical CTM is therefore eliminated right away. However, a more recent constructive version of CTM accepts the contribution of additional information absent from the original encoding. Since the new content about the belt and silver buckle does not disturb the general narrative of the hunting episode, NTM is also considered compatible with our intuitions about content generation. For FTM to qualify an episode as genuine episodic remembering, according to Fernández, it is only necessary that it plays a mnemonic role for the subject (2018: 65). Since the additional information about the belt and silver buckle does not affect the general functionality of that mnemonic role, FTM passes the test as well. Last but not least, STM naturally passes the test due to its constructive character accounting for its conception of episodic memory.

NTM's failure on Fernández's compatibility test is twofold. First, the case of isolated episodes is put forward specifically to challenge NTM. We all have some episodes from our pasts that we are willing to call memories, even though they do not fit into any part of the overall narration of our lives. Unlike the other accounts, NTM's conception of episodic memory relies solely on the fluidity of narration and the sense of storytelling; this makes it unable to accept the kind of memory we have that is detached from our personal past's narrativity. NTM fails here by being too strict. Second, it also fails the compatibility test with our intuition regarding confabulation cases. While the other accounts succeed in ruling out confabulations as genuine episodic memories, NTM does not. Fernández provides the clinical example of Korsakoff's syndrome, a form of amnesia caused by long-term habitual heavy drinking, in which patients tend to mistakenly take imagined events – though described in surprisingly impressive detail – as their personal past experiences. According to NTM, as long as the episode follows the subject's personal storyline, it should be considered a genuine episodic memory. This fails to capture our intuition about confabulation; thus, NTM fails here as well, being too permissive.

Eliminating classical CTM and NTM, the remaining competitors are constructive CTM, STM, and FTM. To single out FTM as the test's sole survivor, Fernández prepares the stage

with the epistemic relevance condition. Recalling an altered, first-person version of Martin and Deutscher's (1966: 167) painter's case, Fernández writes:

Suppose [...] that I am a painter. I decide to paint a bird landing on the roof of a house, and at the end of my work, I have painted that scene in my canvas. However, I am not inclined to think that these are a house and a bird I have ever encountered in the past. I do not think that these are, or were, a real house and a real bird. And yet, it turns out that I am wrong: The house that I have painted is a house that, in fact, I visited as a child. The bird that I have painted is a bird that, in fact, I saw during my visit. And the experience wherein I visualized the scene that I have painted in my canvas causally originates in my past perceptual experience of that bird landing on the roof of that house.

(2018: 57)

Fernández talks about 'epistemic relevance' in the sense that the painter in this case is unaware that his mental representations are episodes of memory rather than imagination. Both classical and constructive CTM straightforwardly count the painter's mental visualization of the house and the bird he painted on the canvas as an episode of remembering, since there is a proper causal connection. Fernández finds this "highly counterintuitive" (2018: 57), for despite the obvious causal connection between the past event and the present visualization, how can we say we remember something when we are unaware of remembering it? STM, on the other hand, even though it doesn't count this visualization as memory in the same way CTMs do, has no reason to deny it as genuine episodic memory. This is because remembering and imagining are part of the same episodic construction system, and episodic remembering in common sense is only one kind of imagining: a subject imagining episodically with more or less reliability. For these reasons, CTMs and STM fail the test at this point. FTM wins the game by being able to cope with our intuition that, in this case, the painter isn't remembering but imagining, because the mental representations put onto the canvas do not play the required mnemonic role that they should if they were to be genuine episodic memories.

Fernández's arguments on how FTM passes the compatibility test are not difficult to follow. This is how he describes the specific function of episodic memory that he calls the mnemonic role: "for any subject *S* and proposition *p*, *S* remembers that *p* just in case *S* has some mental image *i* such that *i* tends to cause in *S* a disposition to believe both that *p* and that *S* experienced that *p*, and *i* tends to be caused in *S* by having experienced that *p*" (2019: 49, emphasis added). Since a genuine instance of episodic remembering needs to play a mnemonic role, as long as an occurring mental representation plays this role – meaning that

6 Identity-Less Memory-Based Personal Identity

it is how memory tends to be produced in the subject's psychological environment – the instance qualifies as episodic memory. If an occurring mental representation provides no indication of such a tendency, it means the mnemonic role isn't present, and therefore the episode isn't an instance of remembering. According to Fernández, there can't be situations where the mnemonic role is at work but the feeling of remembering isn't, or vice versa, because the tendency he mentions includes a personal and subjective feeling of familiarity. However, this does not mean that what the memory content represents (such as the actual experience of a past event) must have occurred for the mental state to count as a memory. While the causal component is necessary as part of the represented content, it is not necessarily required as a metaphysical condition. In other words, a memory can represent itself as causally linked to a past experience, even if no such experience occurred.

This role version of functionalism that Fernández applies to his account of memory seems to rely on the effect it produces in the subject. Memory's function is the functional role required for the subject to feel what-it-is-like to remember (2019: 49). Without further explanation, it seems rational to consider that the mnemonic role is reduced to whenever the subject experiences the feeling of remembering. Then, when Fernández speaks of epistemic relevance in the painter's case, we intuitively consider it a case of imagination, not only because the state in question isn't playing the mnemonic role, but also because the subject experiences no feeling of remembering. In this case, the question of whether the mnemonic role determines the epistemic relevance of an instance is reduced to the question of whether the phenomenology of remembering plays a role in the epistemic relevance of an instance. Related problems include: Is it possible to feel like remembering but in fact not be remembering? And, is it possible to remember without feeling like remembering? I will turn to these questions by the end of the next section.

Furthermore, according to both constructive CTM and STM, the remembering subject is not always in a position to phenomenologically distinguish memory from imagination. For constructive CTM, this is primarily due to its acceptance of content generation, which allows memories to be partly constructed or supplemented beyond what was originally experienced. For STM, the reason is more structural: it sees episodic memory and episodic imagination as products of the same underlying cognitive system – the episodic construction system – based in the hippocampus (Michaelian, 2016). That being said, STM does not deny that memory and imagination can be distinguished in principle (for instance, by appealing to the source

monitoring framework), but it remains agnostic about whether such a distinction is phenomenologically accessible in any given instance. Thus, in cases like the painter example, STM remains neutral as to whether the mental representation is one of memory or imagination. In contrast, constructive CTM, although it allows content generation, still upholds the classical CTM requirement of a causal connection to a past experience. As a result, it places stronger emphasis on distinguishing genuine memory from imagination, maintaining the relevance of epistemic criteria such as those proposed by Fernández.

However, although STM considers remembering as one kind of imagining, this does not mean that remembering is indistinguishable from imagining according to STM. In fact, STM accounts for the “mechanisms responsible for ensuring the reliability of simulational remembering” (Michaelian, 2016: 123).

If STM does not accept epistemic relevance as a necessary condition for episodic remembering, Fernández’s choice to include it has little power in eliminating constructive CTM and STM. We have also seen that STM has a stronger claim to the irrelevance of epistemic relevance than constructive CTM. If STM is correct, the epistemic relevance condition will turn out to be irrelevant for qualifying episodic remembering, which would leave FTM tied with both constructive CTM and STM according to Fernández’s analysis. In that case, he will need additional arguments to single out his champion. In the next section, I will argue that making judgments and providing beliefs have little to do with distinguishing memory from imagination, and that Fernández’s identification of the mnemonic role with the phenomenology of remembering does little to establish epistemic relevance anyway.

Table 3 Summary of the compatibility test. “X” indicates unaccountable; “O” indicates accountable.

		<i>Content Generation</i>	<i>Isolation of Episodes</i>	<i>Confabulation</i>	<i>Epistemic Relevance</i>
<i>Causality</i>	Classical CTM	X	O	O	X
	Constructive CTM	O	O	O	X
<i>Constructive Causality</i>	NTM	O	X	X	O
	FTM	O	O	O	O
	STM	O	O	O	X

6.2 The relevance of epistemic relevance

Fernández wasn’t the first to use the epistemic relevance criterion against causal theories, Debus (2010) has earlier mentioned it as a special difficulty for CTM, and has argued for the importance of epistemic relevance as a legitimate criterion to pose a new challenge to CTM.

6 Identity-Less Memory-Based Personal Identity

To qualify as an instance of episodic memory, Martin and Deutscher (1966) formulated a list of three conditions they consider individually necessary and jointly sufficient, including:

1. Within certain limits of accuracy [the subject] represents that past thing.
2. If the thing was “public,” then [the subject] observed what he now represents. If the thing was “private,” then it was his.
3. [The subject’s] past experience of the thing was operative in producing a state or successive state in him finally operative in producing his representation.

(1966: 166)

The causal connection condition mentioned in the last point characterizes CTM, as well as other CTM-based accounts, including constructive CTM and neoclassical variants. However, in her 2010 paper, where she formulates an epistemic-causal hybrid theory, Debus argues that Martin and Deutscher’s list of necessary conditions must include epistemic relevance in order to be jointly sufficient. She defines this additional condition as follows:

(Epistemic-Relevance-Condition) in order for an experience to count as an instance of remembering, it is necessary that the relevant experience have epistemic relevance for the subject when judging about the past; that is, it is necessary that the subject be disposed to take the irrelevant experience into account when judging about the past.

(Debus, 2010: 21)

To better illustrate the meaning of this condition and to offer good reasons for readers to accept it, Debus compares two individuals, both of whom have “a very normal and average psychological life” (2010: 21), but who differ significantly in their epistemological use of their mental representations. The anomalous individual is John, who is never “disposed to make any epistemological use of the relevant experiences, and he is unaware of the fact that he has reason to take the relevant events into account when judging about the past” (2010: 21). Mary, by contrast, represents normal individuals: she “will make use of her experiential memories” and “is able to form reasonable new beliefs about the past” (2010: 22) on their basis. That is, John is either someone who cannot experience the feeling of remembering – or the phenomenology of memory – that provides him with justification for forming beliefs and making judgments based on his current mental representations, or he is aware that he is remembering but has other reasons for not endorsing any belief or forming any judgment on that basis.

This second case is accepted by Debus as an instance of episodic memory, because there are individuals who are aware of their own tendencies to misremember and thus have low confidence in their beliefs about past experiences, which leads them to deliberately choose not to endorse any belief or form any judgment based on what they think they are remembering (2010: 23). However, Debus does not allow John's mental representations – according to the first interpretation – to count as instances of episodic memory, since they are “not even disposed to make epistemological use of the relevant experience” (2010: 23).

What is missing from Fernández's painter is the same thing that is missing from John: the characteristic feeling of remembering. That is why, even though they have mental representations that causally result from their own past experiences, they are not aware that these representations have the power to justify beliefs that could turn into knowledge. Let us keep in mind that this is the case only if we adopt the traditional framework of CTM, where we choose to distinguish remembering from imagining. However, thinking that identifying memory as memory provides the subject with more epistemic confidence leads us to question the necessity and limited existence of phenomenology in episodic remembering. Does this mean that the feeling of remembering is not actually characteristic of remembering itself, but rather a phenomenology that is distributed quite arbitrarily? And does this imply that STM considers memory as unreliable a faculty as imagination?

Considering both episodic memory and imagination as products of the same episodic construction system does not mean that STM collapses the distinction between the two. While the system can produce a range of representations – including confabulations and imaginings – only those that meet certain reliability conditions, such as an appropriate causal or informational link to the subject's past, qualify as episodic memories. On this view, episodic memory is not merely a more reliable version of imagination, but a distinct kind of output defined in part by its epistemic grounding (Michaelian, 2016). In fact, Michaelian argues that when additional or altered information is integrated during the reconstructive process, the resulting memory can sometimes exhibit *greater* accuracy than the original encoding. On the other hand, Klein and Nichols's (2012) article on patient R.B.'s case provides reasons to believe that at least one aspect of the phenomenology of remembering – namely, the feeling of mineness – is not required for genuine episodic remembering. Furthermore, nonbelieved memories, including later revealed borrowed memories, are known to evoke characteristic feelings of remembering in subjects entertaining those mental representations, even though

these are not veridical instances of episodic memory concerning the subject's own past experiences (Mazzoni et al., 2010). These considerations support the view that although the feeling of remembering generally serves as an indicator of reliable memory, it is not a necessary feature of episodic remembering. In other words, since there are cases in which the feeling of remembering is present for representations of events that did not happen to the subject, and absent for events that did, the phenomenology of remembering appears insufficient as a criterion for genuine memory. If Fernández's analysis of the mnemonic role from the last section is correct, then not only does epistemic relevance lose its status as a phenomenological marker, but functionalism will also need to provide a more refined account of the mnemonic role and its explanatory function.

The claim here is not that epistemic relevance plays no role in episodic remembering, but rather that it should not be treated as a defining criterion. A more robust starting point for defining episodic memory can be found in STM, which understands episodic remembering as the product of a reconstructive process carried out by the episodic construction system: a cognitive system shared with episodic imagination (Michaelian, 2016). According to this view, what qualifies a mental state as an episodic memory is not its phenomenology nor its truthfulness, but its origin in this dedicated system whose function is to simulate past events. This framework allows us to explain why we can remember inaccurately or even falsely, while still relying on a generally reliable system. It also makes sense of cases where subjects feel ownership over imagined or borrowed episodes, and where genuine memories lack a sense of mineness. From this perspective, the simulation system remains reliable on the whole, even if its individual outputs do not always generate epistemically justified beliefs. This model thus provides a flexible but principled account of episodic memory, capable of accommodating both the variability of our subjective experience and the need for epistemic trust in memory as a cognitive faculty.

In the last section of her article, Debus addresses possible objections to accepting her Epistemic-Relevance Condition and considers the reasoning of her "stubborn opponent". She claims that, since proponents of this condition focus on the dissimilarities between Mary and John's mental states, if the opponents instead insist on giving more weight to the similarities between the two mental states, "it seems that there is not much we can do to stop her from doing so" (2010: 24). However, she argues, the stubborn opponent will at least accept that epistemic relevance is a necessary condition for "paradigmatic cases of remembering" (Debus,

2010: 24). Trying to defend Debus's stubborn opponent's position, the relevant question at this stage seems no longer to be the distinction between imagining and remembering, but rather the distinction between remembering and "paradigmatic" remembering.

Looking back at Debus's description of John's peculiar mental state, which seems odd to be considered episodic memory in the common sense, this is quite a radical case where a person has only memories that provide no epistemic relevance. How is such a person supposed to have "a very normal and average psychological life" (2010: 21), as Debus introduces? The fact is that, even if we might find clinical cases with patients having only epistemically irrelevant memories, such cases are not paradigmatic but abnormal, and should not be given too much weight when seeking a general definition of the nature of episodic memory.

In reality, for most individuals in daily life, having present mental representations of past experiences often serves purposes other than seeking accurate information to voluntarily build factual beliefs. That is, even though memory reliability underlies our generation of beliefs and judgments, epistemic relevance does not seem necessary for us to mentally entertain representations of past events that include varying degrees of construction, which we may or may not be comfortable calling instances of memory. Therefore, within the theoretical framework of STM, since our mental representations can be either epistemically relevant or irrelevant, the epistemic relevance condition itself appears to be irrelevant for qualifying memory.

There is a reason why, within the theoretical framework of CTM, Debus regards epistemic relevance as a necessary condition: it stems from the premise that episodic remembering must be distinctly separated from other mental states, most notably from imagining. Accordingly, Debus argues that, to clearly differentiate these phenomena, the single-directional causal relation condition ensures that only epistemically relevant instances qualify as episodic memory, whereas none of Martin and Deutscher's original conditions suffice to exclude properly caused but epistemically irrelevant instances from being classified as episodic remembering. It is against this backdrop that Debus advocates for the inclusion of epistemic relevance as an additional necessary condition.

By contrast, the STM framework attenuates the significance of this distinction. Its primary concern lies not with epistemic relevance per se, but with the overall reliability of episodic memory conceived as a form of episodic imagination generated by a more general

episodic construction system. The fact that Fernández invokes epistemic relevance indicates that he, like proponents of CTM, presupposes a clear ontological and functional distinction between episodic remembering and imagining. Nevertheless, his endorsement of the FTM appears undermined by the recognition that epistemic relevance holds pertinence exclusively within the context of CTM and related memory theories.

6.3 A Simulationist Memory-Based Perdurantist Personal Identity Framework

In this section, I outline the framework of a simulationist, memory-based theory of personal identity. To this end, I revisit the four-dimensional visual representation introduced in Chapter 1. Building on the distinction between three- and four-dimensional perspectives, I further aim to differentiate *subjective* from *objective* conceptions of time.

Take Figure 2: the continuous rectangular prism represents an infinite number of possible events in the universe concerning the lifetime of a person P. Each of the square slices represents a time slice: if slice c is the present time slice for P, slice b represents the moment P comes into existence, slice d is a future instant that P will reach in time, and slice a corresponds to a moment in the objective past when P did not yet exist. The arrow passing through all the slices represents the incessant flow of objective time. In the diagram, the width and height represent the space of all possible worlds, while the depth corresponds to the axis of time. The physical space of movements, as included in one of the dimensions in perdurantism, is not explicitly represented in this diagram; rather, it is embedded in the infinite number of points forming the time slices, each of which represents an event occurring within a spatial environment.

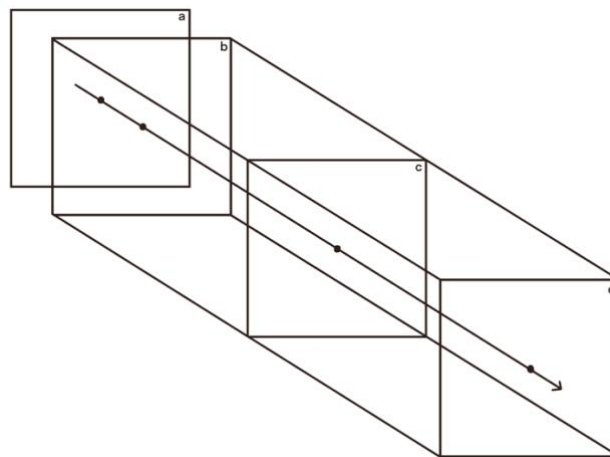


Figure 2 Representation of personal identity in four dimensions.

There are two ways of understanding ‘episodicity’ within the mental time travel framework: (1) when the agent mentally entertains a representation from a first-person perspective and in an experiential manner, the agent can be said to think episodically; and (2) when the mental representations entertained by the agent correspond factually to what happened or what might happen, the agent episodically remembers or imagines. For example, Michaelian (2016: 119) describes how STM considers episodic remembering as a form of episodic imagining. He claims that “I can episodically imagine my great-grandfather’s voyage” from the Ottoman Empire to the United States, which is very similar to his episodically remembering his own voyage from France to Turkey. In this case, episodicity is understood in the first sense: the representational contents when Michaelian imagines being his grandfather, or in his grandfather’s position on the trip, are very similar to those when he remembers being himself on his own trip. He entertains both representations from a first-person perspective and in an experiential manner.

However, to better illustrate the symmetry between remembering and imagining, there is a second way of understanding episodicity. For instance, when I remember having biked to campus yesterday or imagine biking there tomorrow, I am thinking episodically; but when I remember having flown yesterday or imagine flying tomorrow to campus on the back of my dragon, I am not thinking episodically. Rather, I am misremembering, confabulating, misimagining, or voluntarily daydreaming. This is because the mental representations entertained are too far removed from my personal reality in the present world. I will refer to the first understanding as ‘experiential episodicity’ and the second as ‘this-world episodicity’. In what follows, unless otherwise specified, I will adopt the second sense (the sense of ‘this-world’) of episodicity.

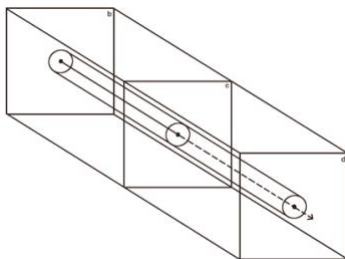


Figure 3 Episodic past, present, and future in four dimensions.

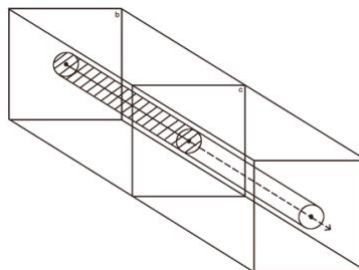


Figure 4 Range of episodic remembering.

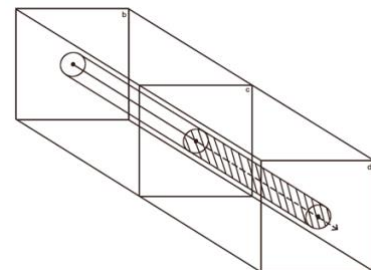


Figure 5 Range of episodic imagining.

If the rectangular prism represents the infinite number of experiential events across P’s possible worlds, then the tube in the middle corresponds to P’s range of episodic

6 Identity-Less Memory-Based Personal Identity

thinking.²² The segment of the tube extending from slice b to c can be identified as episodic remembering, while the segment from slice c to d corresponds to episodic imagining. The only difference between the two is the direction in which the thinking is oriented. To avoid terminological confusion, although the simulationist framework conceptualizes episodic remembering as reliable, past-oriented episodic imagining, I will use the term ‘episodic remembering’ to refer to past-oriented episodic imagining and ‘episodic imagining’ to refer to future-oriented episodic imagining.²³ If the penetrating arrow represents the actual succession of events experienced by p over time, then all points within the episodic tube fall within the acceptable limits of reconstruction and generation of representational content, such that the mental states involved remain genuine instances of episodic thinking.

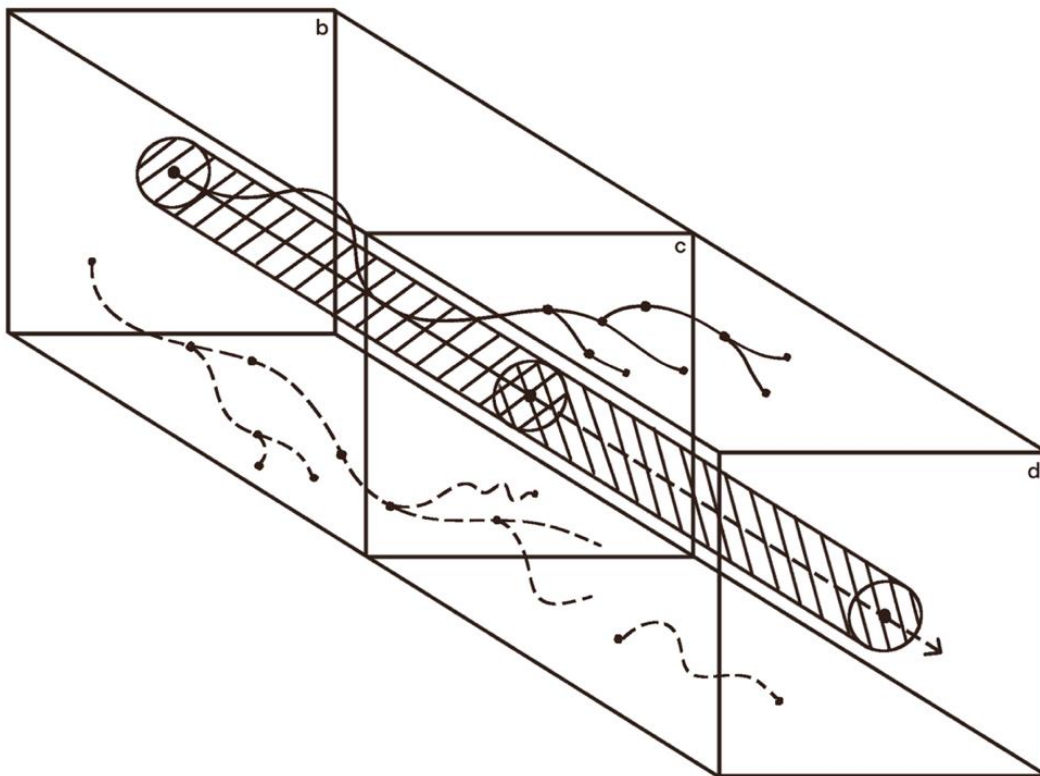


Figure 6 Psychological activities in four dimensions: episodic mental time travel and pure daydream.

²² The exact width of the tube is open to debate. That is, the limits of the reconstruction process and the extent of generative content required for instances to qualify as episodic thinking are beyond the scope of this work. Moreover, it is uncertain whether such a precise delineation is even feasible, given the inherent instability of human thought processes.

²³ There is a slight difference between ‘past-oriented episodic imagining’ and what we might call ‘past-oriented counterfactual thinking’: considering them in terms of events with durations – that is, having a temporal characteristic – past-oriented episodic imagining includes certain points on time-slices that lie inside the episodic tube, whereas past-oriented counterfactual thinking has the entire timeline of the event fully outside the tube. Considering only isolated time slices amounts to examining moments outside their context, which is unlikely to yield fruitful insights for a perdurantist time travel framework.

When we mentally entertain representations of experiences in the first person, there are several possible variations. First, the most rigorous form is strict episodic remembering, where we recall a sequence from an episodic point on slice *b* to another episodic point on slice *c*, and the line connecting these two points remains entirely within the episodic tube from slice *b* to *c*. Second, I may be episodically remembering but in a more fluid way, without paying strict attention to what actually happened, causing me to slip outside the episodic tube briefly before returning. I could also choose not to return to the tube: this could be a situation where I am episodically remembering but suddenly begin to wonder what would have happened if I had made different choices that would have drastically altered my episodic tube. This often occurs not when we intend to remember an episode for a specific purpose, but when we are prompted to think about a past experience and allow our mind to wander. If this episodic thought gets interrupted once it leaves the episodic tube, it may also stop outside the tube without returning. Regardless of whether the thought returns to the episodic tube, this kind of daydreaming is fact-based: the starting point is episodic, and the development toward the non-episodic point is also episodic. We might say this type of daydreaming is episodically based. Third, there are times when we imagine things about ourselves that are far from factual. I can daydream about the past, such as imagining how my childhood would have been if I had two younger sisters instead of brothers (outside but close to the episodic tube), or if I had two ogres as siblings instead of younger brothers (outside and farther from the episodic tube). Similarly, I can daydream about the future, wondering what would happen if tomorrow I don't bike to campus but buy a car (outside but close to the episodic tube), or train a dragon to fly me there (outside and farther from the episodic tube).²⁴

These are the three kinds of first-person thought distinguished in this framework, corresponding to three levels of episodocity: episodic thinking, episodic-based daydreaming, and daydreaming. This way of understanding episodocity is compatible with the nature of the feeling of mineness in episodic remembering, as well as with what we understand about memory through the concept of quasi-remembering.

²⁴ Some might argue that buying a car is sufficiently close to an individual's personal reality and is conceivable enough to qualify as episodically based. However, engaging in this discussion involves determining the boundaries of the episodic tube. The advantage of this framework lies in its ability to represent the concept of episodic thinking without requiring a precise empirical or conceptual definition of these boundaries. Indeed, I can episodically entertain representations that appear quite distant from my personal reality, while the thought remains within the episodic tube.

6 Identity-Less Memory-Based Personal Identity

When I think of a past event in the first person and in an experiential way, the mental representation can lie within the episodic tube without my endorsing the feeling that it is my own past event. It is equally possible that I mentally entertain a past-oriented daydream situated outside the episodic tube while still feeling as if I own that experience.²⁵ The feeling of ownership does not delineate the boundary between inside and outside the episodic tube. Several factors determine whether a mental state qualifies as episodic thinking: the reliability of an individual's metacognitive system decides whether the person's episodic tube boundary is traced in such a way that positive information about the person can be obtained; a person's psychological continuity can only provide positive information about them if there is coherence among the points. If such coherence exists, the points lie within the episodic tube and correspond to episodic points (i.e., episodic mental representations). This means that simulationist memory is reliable in healthy individuals, who base their personal identity on simulationist mental representations that construct a comprehensible personal narrative.

Similarly, a quasi-memory that was not experienced by the rememberer will be situated outside the episodic tube. This view aligns with the idea that when fusion and fission occur, the original persons involved do not survive. For example, after fission, what I remember in the first person about the experiences of the original person will be excluded from the episodic tube, regardless of whether I phenomenologically feel any ownership over those memories. The same applies to fusion: when two lines of psychological continuity merge into one, the original individuals cease to exist and a new person is formed. Even if the new individual quasi-remembers two distinct past experiences from the same past time slice, these will be two points on that time slice lying outside the new individual's episodic tube.²⁶

This framework aims to capture personal identity through the episodic tube, which is composed of mental representations classified as episodic remembering and imagining within the simulationist framework. It is compatible with a version of the hybrid theory of personal

²⁵ There could be multiple reasons for this. It may be because the entertained representation is situated close to the episodic tube, and its similarity to the episodic range of the actual event makes it feel familiar rather than unfamiliar. Alternatively, even if the representation lies far from the episodic tube, it could correspond to a point on that time slice that I frequently revisit. Through the mechanism of repeatedly revisiting this daydream, it gradually acquires a sense of ownership.

²⁶ It will lie outside the episodic tube in two senses: first, because the two original individuals are ancestors of the new individual, the two quasi-remembered points will be located on a past time slice of objective time – before the new individual has come into existence; second, even if one assumes that the episodic tube of the new individual existed prior to fusion, the two simultaneous points on that time slice cannot both lie inside the tube.

identity, insofar as simulationist memory requires a metacognitive system that depends on bodily components such as the hippocampus for its production. Without such bodily elements, the episodic points constituting the episodic tube cannot be produced, and thus no person exists. Importantly, this neo-Lockean framework of personal identity is not circular, as simulationist memory does not presuppose personal identity.

6.4 Conclusion

In this final chapter of my thesis, I discussed Fernández's (2019) compatibility test with the aim of distinguishing the STM from rival theories of memory, namely, the classical and constructive CTM, the NTM, and an additional variant of the STM. One particular criterion used during the test – epistemic relevance – raised important questions regarding its pertinence. In analyzing a more detailed formulation of the Epistemic-Relevance Condition proposed by Debus, I concluded that epistemic relevance is irrelevant for formulating a metaphysical theory of episodic remembering within the STM framework. STM appears to withstand the critique that it blurs the boundary between memory and imagination by treating them as products of the same episodic construction system by appealing to the source monitoring framework to account for their distinction.

Drawing on a perdurantist visual representation of personal identity, I further argued that STM is the most compatible theory with the nature of memory as constitutive of personal identity. Notably, STM's complete rejection of the causal connection condition liberates quasi-memory from the strict requirement that one can only remember one's own past experiences. Phenomenological arguments supporting this position align more closely with the theoretical framework of STM than with CTM or other CTM-based accounts.

What is particularly fruitful in the episodic tube framework proposed here is that it integrates three distinct theoretical dimensions into a unified model: (1) a perdurantist metaphysics of persistence through time, (2) the simulationist understanding of episodic memory and imagination as products of a shared constructive system, and (3) a bounded notion of factivity that preserves psychological continuity without requiring strict causal connections. The tube metaphor, though illustrative, offers conceptual precision by demarcating a zone of episodic thought that is reconstructive yet constrained: where memory remains sufficiently anchored to the subject's temporal self without relying on infallible retrieval. This, I argue, provides a new approach to the circularity problem traditionally posed

6 Identity-Less Memory-Based Personal Identity

to neo-Lockean theories: by shifting focus from factive memory to the broader function of episodic construction, we can account for psychological continuity without presupposing identity. In this view, the episodic tube consists of person-stages (time slices) unified not by body or biological sameness, but by the cognitive ability to mentally travel backward and forward across one's own timeline. STM, framed this way, avoids the circular dependence on identity that undermines classical memory theories and instead grounds personal identity in the ongoing functional coherence of episodic thought.

Conclusion

This thesis was motivated by a gap in the literature between memory theory and personal identity theory. As causal theories of memory evolve into more constructive or simulation-based models, and as post-causal theories gain popularity, neo-Lockean approaches to personal identity have yet to engage meaningfully with simulationist memory. This work set out to examine whether a simulation theory of memory could support a version of neo-Lockean personal identity without falling into circularity. The working hypothesis was that, if memory is to constitute personal identity without presupposing it, then memory must be identity-less: that is, it must not rely on prior personal identity to function.

My approach was tripartite: I laid the conceptual groundwork, examined the inadequacy of current strategies for avoiding circularity, and finally proposed a framework in which memory can ground personal identity in a non-circular way.

The first part of the thesis examined the metaphysics of persistence. I argued for a genuine philosophical difference between endurantism and perdurantism in how they account for diachronic identity. Perdurantism, with its four-dimensional ontology, better accommodates identity over time by allowing for spatial and temporal mereological changes. This identity-less framework for objects provides a more promising background for thinking about persons. I then considered which accounts of personal identity are most compatible with this framework. Mind-based, body-based, and hybrid views can all be adapted to perdurantism, but a hybrid neo-Lockean view – one that allows for minimal bodily continuity – was selected for its balance between metaphysical rigor and psychological realism. The challenge then became how to avoid circularity in this model, given that memory often seems to presuppose identity.

In the second part, I investigated the root of the circularity problem by analyzing the historical development of memory theories and their attempts to respond to this objection. I concluded that post-causal accounts – those that drop the requirement for a causal connection between the rememberer and the remembered experience – are most promising. Before committing to any specific theory, I analyzed whether identity-less memory is

Conclusion

conceptually possible. Drawing from linguistic analysis, particularly Anscombe's account of the object-less use of 'I', I argued that the use of the first-person pronoun in memory claims does not guarantee identity. This helped support the idea that quasi-memory can be used without presupposing personal identity.

I then extended this insight from language to phenomenology. The contingency of the feeling of mineness – that is, the subjective sense that a memory is one's own – shows that memory can operate without necessarily grounding identity. Thus, both linguistically and metaphysically, memory can be identity-less. This conception diverges significantly from traditional views that regard memory as a reliable source of self-knowledge. In the final chapter, I examined various memory theories through the lens of epistemic relevance, ultimately singling out the simulation theory of memory as the best suited for a neo-Lockean framework that understands persons as perduring, four-dimensional entities.

This thesis aims to contribute to the philosophical literature by bridging memory theory and personal identity through a novel combination of perdurantism, simulationism, and neo-Lockeanism. It invites further inquiry into how simulationist memory could inform legal, psychological, or ethical discussions about responsibility and selfhood. While this work remains largely metaphysical – though not without some scientific grounding – future research could explore empirical or computational models of identity-less memory, thereby deepening the connection between cognitive science and metaphysical theorizing. Multiple future directions remain open.

There are, indeed, several aspects I have not had the time or space to develop fully in this thesis. The episodic tube framework, for instance, aims to embody the virtues of perdurantist persistence and simulation-based personal identity. However, since its delimitation cannot yet be precisely defined, the notion of the tube remains, for now, more metaphorical than systematically explanatory. Given that the central conclusion of this thesis is that a non-circular neo-Lockean view requires a minimal bodily criterion, the episodic tube framework must eventually be able to accommodate not only one psychological continuity but also multiple-occupancy scenarios such as fusion, fission, and reduplication. Furthermore, the metaphysical status of the tube remains unclear. Through a perdurantist lens, I proposed an identity-less account of persistence over time. But if identity over time is not underwritten by a metaphysical self, what exactly is it that the tube represents?

Identity-Less Personal Past

Regarding the role of the simulation theory of memory, I regret not being able to delve more deeply into its precise relation to personal identity. At this stage, it remains ambiguous whether simulationist memory actually *constitutes* personal identity or merely *tracks* it. This ambiguity is directly connected to another unresolved issue: the status of the bodily criterion. Future work will need to clarify the respective metaphysical roles of the body and the mind in personal identity. While I have claimed that the mind is fundamental, I lean toward the view that the mind constitutes personal identity, whereas the body is only partially necessary, serving as a causal interface with the world. A more thorough analysis is required to ground this position.

Finally, the notion of identity-less-ness itself is philosophically provocative. It challenges our intuitive beliefs about ethical, moral, and legal responsibility over time. If the contingency of the feeling of mineness indeed reveals a form of identity-less memory, then how can we be held accountable for past actions? Although I have argued for a form of identity-less identity – where identity is conceived as something continuously constructed and reconstructed – I have not been able to explore its ethical implications in depth. I suspect that these implications are not insoluble, but they certainly call for further investigation.

In addition to these open questions, my inquiry gives rise to several further lines of research: If identity-less memory can ground a robust theory of personal identity, how do we practically distinguish remembering from imagining? Can we develop a more precise epistemology of simulationist remembering? What would the criteria of identity look like if both ownership and causal origin are contingent? If perduring persons are constructed from episodic tubes, where exactly do these tubes begin or end – and can such boundaries ever be clearly defined? Whether these questions can be answered remains to be seen. I leave the weight of these inquiries to future investigations.

Bibliography

- American Psychiatric Association. (2013). *DSM-5: diagnostic and statistical manual of mental disorders* (5th edition). American Psychiatric Publishing.
- Anscombe, G. E. M. (1975). The first person. In S. Guttenplan (Ed.), *Mind and language*, 45–65. Oxford University Press.
- Anscombe, G. E. M. (2000). *Intention*. Harvard University Press.
- Bailey, A. (2015). Animalism. *Philosophy Compass*, 10(12), 867–883.
<https://doi.org/10.1111/phc3.12263>
- Bernecker, S. (2008). *The metaphysics of memory*. Springer.
- Bernecker, S. (2010). *Memory: A philosophical study*. Oxford University Press.
- Butler, J. (2008). Of Personal Identity. In J. Perry (Ed.), *Personal Identity* (2nd ed.), 99–105. University of California Press.
- Campbell, L. (2018). An epistemology for practical knowledge. *Canadian Journal of Philosophy*, 48(2), 159–177. <https://www.jstor.org/stable/10.2307/48547306>
- Copenhaver, R. (2018). Reid on Memory and Personal Identity. *The Stanford Encyclopedia of Philosophy* (Edward N. Zalta ed.).
<https://plato.stanford.edu/archives/win2018/entries/reid-memory-identity/>
- Coliva, A. (2003). The First Person: Error through Misidentification, the Split between Speaker's and Semantic Reference, and the Real Guarantee. *The Journal of Philosophy*, 100(8), 416–431. <https://www.jstor.org/stable/3655580>
- Coliva, A. (2006). Error through Misidentification: Some Varieties. *The Journal of Philosophy*, 103(8), 403–425. <https://www.jstor.org/stable/20619956>
- Coliva, A. & Palmira, M. (2024). Immunity to error through misidentification: some trends. *Philosophical Psychology*, 38(3).
<https://doi.org/10.1080/09515089.2024.2371423>
- Coliva, A. & Palmira, M. (2025). Errors through misidentification and the specialness of first-person thought. *Philosophical Psychology*.
<https://doi.org/10.1080/09515089.2025.2452766>
- Debus, D. (2010). Accounting for Epistemic Relevance: A New Problem for the Causal Theory of Memory. *American Philosophical Quarterly*, 47(1), 17–29.
<https://www.jstor.org/stable/40606900>
- De Brigard, F. (2014). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*, 191(2), 155–185.
<https://doi.org/10.1007/s11229-013-024-7>
- De Brigard, F. (2023). *Memory and Remembering*. Cambridge University Press.
<https://doi.org/10.1017/9781108955447>
- Dings, R. & McCarroll, C. R. (2022). The Complex Phenomenology of Episodic Memory: Felt Connections, Multimodal Perspectivity, and Multifaceted Selves. *Journal of Consciousness Studies*, 29(11-12), 29–55.
<https://doi.org/10.53765/20512201.29.11.029>

- Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Harvard University Press.
- Donnelly, M. (2011). Endurantist and perdurantist accounts of persistence. *Philosophical Studies*, 154(1), 27–51. <https://www.jstor.org/stable/41487652>
- Doyle, J. (2016). ‘Spurious egocentricity’ and the first person. *Synthese*, 193(11), 3579–3589. <https://www.jstor.org/stable/10.2307/48726363>
- D’Argembeau, A. (2012). Autobiographical memory and future thinking. In D. Berntsen, D. C. Rubin (Eds.), *Understanding Autobiographical Memory: Theories and Approaches*, 311–330. Cambridge University Press.
- Elvevåg, B., Fisher, J. E., Weickert, T. W., Weinberger, D. R., & Goldberg, T. E. (2004). Lack of false recognition in schizophrenia: a consequence of poor memory? *Neuropsychologia*, 42, 546–554. <https://doi.org/10.1016/j.neuropsychologia.2003.08.013>
- Evans, G. (1982). *The Varieties of Reference*. Oxford University Press.
- Fernández, J. (2018). The Functional Character of Memory. In K. Michaelian, D. Debus & D. Perrin (Eds.), *New Directions in the Philosophy of Memory*, 52–72, Routledge.
- Fernández, J. (2019). *Memory: A Self-Referential Account*. Oxford University Press.
- Frege, G. (1948). Sense and Reference. *The Philosophical Review*, 57(3), 209–230. <https://doi.org/10.2307/2181485>
- García-Carpintero, M. (2024). Memory-based reference and immunity to error through misidentification. *Synthese*, 204(70). <https://doi.org/10.1007/s11229-024-04664-2>
- Garret, B. (1994). Anscombe and The First Person. *Crítica: Revista Hispanoamericana de Filosofía*, 26(78), 97–113. <http://www.jstor.org/stable/40104714>
- Garret, B. (1997). Anscombe On ‘I’. *The Philosophical Quarterly (1950-)*, 47(189), 507–511. <https://www.jstor.org/stable/2956282>
- Gordon-Roth, J. (2023). Locke on Personal Identity. *The Stanford Encyclopedia of Philosophy* (Edward N. Zalta & Uri Nodelman eds.). <https://plato.stanford.edu/archives/fall2023/entries/locke-personal-identity/>
- Grice, H. P. (1941). Personal Identity. *Mind*, 50(200), 330–350. <http://www.jstor.org/stable/2250897>
- Haddock, A. (2019). “I am NN”: A Reconstruction of Anscombe’s “The First Person”. *European Journal of Philosophy*, 27(4), 957–970. <https://doi-org.sid2nomade-2.grenet.fr/10.1111/ejop.12445>
- Halder, S., Juel, B. E., Nilsen, A. S., Raghavan, L. V., & Storm, J. F. (2021). Changes in measures of consciousness during anaesthesia of one hemisphere (Wada test). *NeuroImage*, 226, 117566. <https://doi.org/10.1016/j.neuroimage.2020.117566>
- Hawley K. (2008). Persistence and Determination. *Royal Institute of Philosophy Supplement*, 62, 197–212. <https://doi.org/10.1017/S1358246108000647>
- Hume, D. (2008). Of Personal Identity. In J. Perry (Ed.), *Personal Identity* (2nd ed.), 161–172. University of California Press.
- Johnson, M. (2010). *Surviving Death*. Princeton University Press.
- Kind, A. (2020). *Philosophy of Mind: The Basics*. Routledge.
- Klein, S. B. (2014). Auto-noesis and belief in a personal past: An evolutionary theory of episodic memory indices. *Review of philosophy and Psychology*, 5(3), 427–447. <https://doi.org/10.1007/s13164-014-0181-8>

Bibliography

- Klein, S. B. (2015) What memory is. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(1), 1–38. <https://doi.org/10.1002/wcs.1333>
- Klein, S. B. & Nichols, S. (2012). Memory and the sense of personal identity. *Mind*, 121(483), 677–702. <https://doi.org/10.1093/mind/fzs080>
- Klein, S. B., Loftus, J., & Kihlstrom, J. F. (1996). Self-knowledge of an amnesic patient: toward a neuropsychology of personality and social psychology. *Journal of Experimental Psychology: General*, 125(3), 250–260. <https://doi.org/10.1037//0096-3445.125.3.250>
- Klein, S. B., Loftus, J., & Kihlstrom, J. F. (2002). Memory and temporal experience: the effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. *Social Cognition*, 20(5), 353–379. <https://doi.org/10.1521/soco.20.5.353.21125>
- Kolibius, L., Roux, F., Parish, G., Ter Wal, M., Van Der Plas, M., Chelvarajah, R., Sawlani, V., Rollings, D., Lang, J., Gollwitzer, S., Walther, K., Hopfenngärtner, R., Keiselmeyer, G., Hamer, H., Staresina, B., Wimber, M., Bowman, H. & Hanslmayr, S. (2023). Hippocampal neurons code individual episodic memories in humans. *Nature Human Behaviour*. *Nature Human Behaviour*, 7, 1968–1979. <https://doi.org/10.1101/2021.06.28.450149>
- Langland-Hassan, P. (2022). Propping up the causal theory. *Synthese*, 200(95). <https://doi.org/sid2nomade-1.grenet.fr/10.1007/s11229-022-03635-9>
- Lewis, D. (1983). Survival and Identity: Postscript. In D. Lewis, *Philosophical Papers*, . Oxford University Press.
- Lewis, D. (1986). *On the Plurality of Worlds*. Blackwell.
- Locke, J. (2008). In J. Perry (Ed.), *Personal Identity* (2nd ed.), 33–52. University of California Press.
- Locke, J. (2014). *An Essay Concerning Human Understanding*. Wordsworth Editions Limited.
- Lucci, D. (2021). The Biblical Roots of Locke's Theory of Personal Identity. *Zygon: Journal of Religion and Science*, 56(1), 168–187. <https://doi.org/10.1111/zygo.12674>
- Madden, R. (2017). Animal Self-Awareness. *Philosophy, Theory, and Practice in Biology*, 9(9), 1–14. <https://doi.org/10.3998/ptb.6959004.0009.009>
- Madden, R. (2024). *Personal Identity and the Self*. Cambridge University Press.
- Magidor, O. (2016). Endurantism vs. Perdurantism?: A Debate Reconsidered. *Noûs*, 50(3), 509–532. <https://www.jstor.org/stable/26631403>
- Mammarella, N., Altamura, M., Padalino, F. A., Petito, A., Fairfield, B., & Bellomo, A. (2010). False memories in schizophrenia? An imagination inflation study. *Psychiatry Research*, 179, 267–273. <https://doi.org/10.1016/j.psychres.2009.05.005>
- Martin, C. B. & Deutscher, M. (1966). Remembering. *The Philosophical Review*, 75(2), 161–196. <https://doi.org/10.2307/2183082>
- Mazzoni, G., Scoboria, A. & Harvey, L. (2010). Nonbelieved Memories. *Psychological Science*, 21(9), 1334-1340. <https://www.jstor.org/stable/41062374>
- McGrath, M. (2007). Four-Dimensionalism and the Puzzles of Coincidence. In D. W. Zimmerman (Ed.), *Oxford Studies in Metaphysics* 3, 143–176. Oxford University Press.
- Michaelian, K. (2011). Generative memory. *Philosophical Psychology*, 24(3), June 2011, 323–342. <https://doi.org/10.1080/09515089.2011.559623>

- Michaelian, K. (2016). *Mental Time Travel: Episodic Memory and Our Knowledge of the Personal Past*. The MIT Press.
- Michaelian, K. (2020). Confabulating as Unreliable Imagining: In Defence of the Simulationist Account of Unsuccessful Remembering. *Topoi*, 39, 133–148.
<https://doi.org/10.1007/s11245-018-9591-z>
- Michaelian, K. (2022). Radicalizing simulationism: Remembering as imagining the (nonpersonal) past. *Philosophical Psychology*, 37(5), 1–27.
<https://doi.org/10.1080/09515089.2022.2082934>
- Michaelian, Kourken & Robins, Sarah (2018). Beyond the causal theory? Fifty years after Martin and Deutscher. In K. Michaelian, D. Debus & D. Perrin (Eds.), *New Directions in the Philosophy of Memory*, 13–32, Routledge.
- Muñoz-Corcuera, A. (2021). The Physical, the Subjective, and the Social in the Debate on Personal Identity: A Critical and Historical Overview of Neo-Lockean Accounts. *Sujetos, personas e individuos: estudios sobre la individualidad*, 28, 22–45.
<https://doi.org/10.15366/bp2021.28.001>
- Nagel, T. (1971). Brain Bisection and the Unity of Consciousness. *Synthese*, 22(3/4), 396–413.
<http://www.jstor.org/stable/20114764>
- Nerlich, G. C. (1959). On evidence for identity. *Australasian Journal of Philosophy*, 37(3), 201–214. <https://doi-org.sid2nomade-1.grenet.fr/10.1080/00048405985200221>
- Noonan, H. (2019). *Personal Identity* (3rd ed.). Routledge.
- Oderberg, D. S. (1993). *The Metaphysics of Identity over Time*. The Macmillan Press.
- Olson, E. (1997). *The Human Animal: Personal Identity without Psychology*. Oxford University Press.
- Olson, E. (2002). Thinking animals and the reference of ‘I’. *Philosophical Topics*, 30(1), 189–207. <https://www.jstor.org/stable/43154385>
- Olson, E. (2003). An argument for animalism. In R. Martin & J. Barresi (Eds.), *Personal Identity*, Blackwell.
- Olson, E. (2023). Personal Identity. *The Stanford Encyclopedia of Philosophy* (Edward N. Zalta & Uri Nodelman eds.).
<https://plato.stanford.edu/archives/fall2023/entries/identity-personal/>
- Otgaar, H., Scoboria, A. & Mazzoni, G. A. L. (2014). On the existence and implications of nonbelieved memories. *Current Directions in Psychological Science*, 23(5), 349–354. <https://doi.org/10.1177/0963721414542102>
- Otgaar, H., Bucken, C., Bogaard, G., Wade, K. A., Hopwood, A. R., Scoboria, A. & Howe, M. L. (2019). Nonbelieved memories in the false memory archive. *Journal of Applied Research in Memory and Cognition*, 8(4), 429–438.
<https://doi.org/10.1016/j.jarmac.2019.07.003>
- Parfit, D. (1971). Personal identity. *The Philosophical Review*, 80(1), 3–27.
<https://doi.org/10.2307/2184309>
- Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.
- Parfit, D. (2012). We Are Not Human Beings. *Philosophy*, 87(339), 5–28.
<http://www.jstor.org/stable/41441489>
- Parsons, J. (2000). Must a Four-dimensionalist Believe in Temporal Parts? *The Monist*, 83(3), 399–418. <http://www.jstor.org/stable/27903694>
- Perrin, D. (2016). Asymmetries in Subjective Time. In K. Michaelian, S. B. Klein, & K. K. Szpunar (Eds.), *Seeing the Future: Theoretical Perspectives on Future-Oriented*

Bibliography

- Mental Time Travel*, Oxford Academic.
<https://doi.org/10.1093/acprof:oso/9780190241537.003.0003>
- Perrin, D. & Michaelian, K. (2017). Memory as mental time travel. In S. Bernecker and K. Michaelian (Eds.), *The Routledge Handbook of Philosophy of Memory*, 169–179. Routledge.
- Perrin, D., Sant’Anna, A. (2022). Episodic memory and the feeling of pastness: from intentionalism to metacognition. *Synthese*, 200(109). <https://doi.org/sid2nomade-2.grenet.fr/10.1007/s11229-022-03567-4>
- Perrin, D., Michaelian, K. & Sant’Anna, A. (2020). The Phenomenology of Remembering Is an Epistemic Feeling. *Frontiers in Psychology*, 11(1531), 1–14.
<https://doi.org/10.3389/fpsyg.2020.01531>
- Perry, J. (2008). *Personal Identity* (2nd ed.). University of California Press.
- Piolino, P., Desgranges, B., Clarys, D., Guillery-Girard, B., Taconnat, L., Isingrini, M. & Eustache, F. (2006). Autobiographical Memory, Autonoetic Consciousness, and Self-Perspective in Aging. *Psychology and Aging*, 21(3), 510–525.
<https://doi.org/10.1037/0882-7974.21.3.510>
- Prior, A. N. (1967). On Spurious Egocentricity. *Philosophy*, 42(162), 326–335.
<https://www.jstor.org/stable/3748495>
- Pryor, J. (1999). Immunity to Error through Misidentification. *Philosophical Topics*, 26(½), 271–304. https://www.jstor.org/stable/43154287?seq=1&cid=pdf-reference#references_tab_contents
- Quine, W. (1961). *From a Logical Point of View* (2nd ed.). Harper Torchbooks.
- Quine, W. (1976). Worlds Away. *The Journal of Philosophy*, 73(22), 859–863.
<https://www.jstor.com/stable/2025761>
- Reid, T. (2008a). Of Identity. In J. Perry (Ed.), *Personal Identity* (2nd ed.), 107–112. University of California Press.
- Reid, T. (2008b). Of Mr. Locke’s Account of Our Personal Identity. In J. Perry (Ed.), *Personal Identity* (2nd ed.), 113–118. University of California Press.
- Rivadulla-Duró, A. (2024). The Simulation Theory of Memory and the phenomenology of remembering. *Phenomenology and the Cognitive Sciences*, 23, 925–945.
<https://doi.org/10.1007/s11097-022-09881-z>
- Roache, R. (2006). A Defence of Quasi-Memory. *Philosophy*, 81(316), 323–355.
<https://www.jstor.org/stable/4127438>
- Roache, R. (2016). Memory and mineness in personal identity. *Philosophical Psychology*, 29(4), 479–489. <http://dx.doi.org/10.1080/09515089.2015.1102216>
- Roux, F., Parish, G., Chelvarajah, R., Rollings, D., Sawlani, V., Hamer, H., Gollwitzer, S., Kreiselmeyer, G., Ter Wal, M., Kolibius, L., Staresina, B., Wimber, M., Self, M. & Hanslmayr, S. (2022). Oscillations support short latency co-firing of neurons during human episodic memory formation. *eLife*, 11(e78109).
<https://doi.org/10.7554/eLife.78109>
- Sant’Anna, A. (2021). Attitudes and the (dis)continuity between memory and imagination. *Estudios de Filosofía*, 64, 73–93. <https://doi.org/10.17533/udea.ef.n64a04>
- Sattig, T. (2018). Memory-based personal identity without circularity. In V. Buonomo (Ed.), *The Persistence of Persons: Studies in the Metaphysics of Personal Identity over Time*, Editiones Scholasticae.
- Schechtman, M. (1990). Personhood and Personal Identity. *The Journal of Philosophy*, 87(2), 71–92. <https://www.jstor.org/stable/2026882>

- Schechtman, M. (1996). *The Constitution of Selves*. Cornell University Press.
- Schechtman, M. (2011a). Memory and identity. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 153(1), 65–79.
- Schechtman, M. (2011b). The Narrative Self. In Shaun Gallagher (Ed.), *The Oxford handbook of the self*, 394–416. Oxford University Press.
- Schechtman, M. (2024). Personal identity and mental time travel. *Philosophy and the Mind Sciences*, 5, 1–21. <https://doi.org/10.33735/phimisci.2024.10639>
- Shanton, K. & Goldman, A. (2010). Simulation theory. *WIREs Cognitive Science*, 1, 527–538. <https://doi.org/10.1002/wcs.33>
- Shoemaker, S. (2003a). Persons and their pasts. In S. Shoemaker (Ed.), *Identity, Cause, and Mind* (2nd ed.), 19–48. Oxford University Press.
- Shoemaker, S. (2003b). Self-reference and self-awareness. (Ed.), *Identity, Cause, and Mind* (2nd ed.), 6–18. Oxford University Press.
- Shoemaker, D. (2008). Personal Identity and Memory. In J. Perry (Ed.), *Personal Identity* (2nd ed.), 119–134. University of California Press.
- Shoemaker, S. (2012). Against simplicity. In G. Gasser & M. Stefan (Eds.), *Personal Identity: Complex or Simple?* Cambridge University Press.
- Sider, T. (1997). Four-Dimensionalism. *The Philosophical Review*, 106(2), 197–231. <https://www.jstor.org/stable/2998357>
- Slors, M. (2001). Personal Identity, Memory, and Circularity: An Alternative for Q-Memory. *The Journal of Philosophy*, 98(4), 186–214. <http://www.jstor.org/stable/2678477>
- Snowdon, P. F. (2014). *Persons, Animals and Ourselves*. Oxford University Press.
- Sperry, R. W. (1968). Hemisphere deconnection and unity in conscious awareness. *American Psychologist*, 23(10), 723–733. <https://doi.org/10.1037/h0026839>
- Stainton, R. J. (2019). Re-reading Anscombe on 'I'. *Canadian Journal of Philosophy*, 49(1), 70–93. <https://www.jstor.org/stable/10.2307/48547331>
- Strawson, G. (1999). The Self. In S. Gallagher, & J. Shear (Eds.), *Models of the Self*, 3–45. Imprint Academic.
- Strawson, G. (2014). Circularity? In G. Strawson (Ed.), *Locke on Personal Identity: Consciousness and Concernment*. <https://doi.org/10.23943/princeton/9780691161006.003.0017>
- Strawson, P. F. (1959). *Individuals: An Essay in Descriptive Metaphysics*. Taylor & Francis.
- Sugar, J. & Moser, M.-B. (2019). Episodic memory: Neuronal codes for what, where, and when. *Hippocampus*, 29(12), 1190–1205. <https://doi-org.sid2nomade-1.grenet.fr/10.1002/hipo.23132>
- Sutton, J. (1998). *Philosophy and memory traces: Descartes to connectionism*. Cambridge University Press.
- Swinburne, R. (1976). Persons and personal identity. In H. D. Lewis (Ed.), *Contemporary British Philosophy*, Allen & Unwin.
- Textor, M. (2011). *Routledge Philosophy GuideBook to Frege on Sense and Reference*. Routledge.
- Trakas, M. (2021). The sense of mineness in personal memory: problems for the endorsement model. *Estudios de Filosofía*, 64, 155–172. <https://doi.org/10.17533/udea.ef.n64a08>
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory*. Academic Press.

Bibliography

- Tulving, E. (1985). Memory and Consciousness. *Canadian Psychology / Psychologie canadienne*, 26(1), 1-12. <https://doi.org/10.1037/h0080017>
- Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology*, 53(1), 1–25. <https://doi-org.sid2nomade-2.grenet.fr/10.1146/annurev.psych.53.100901.135114>
- Wasserman, R. (2002). The Standard Objection to the Standard Account. *Philosophical Studies*, 111, 197–216. <https://doi.org/10.1023/A:1021283405720>
- Williams, B. (1973a). *Problems of the Self*. Cambridge University Press.
- Williams, B. (1973b). Are persons bodies? In B. Williams (Ed.), *Problems of the Self*, 64–81. Cambridge University Press.
- Wittgenstein, L. (1958). *Philosophical Investigations* (G. E. M. Anscombe, Trans.). Basil Blackwell.
- Wittgenstein, L. (1958). *The Blue and Brown Books*. Harper & Row.