



HAL
open science

La mémoire constructive : repenser la mémoire pour penser l'identité personnelle

Loraine Gérardin-Laverge

► To cite this version:

Loraine Gérardin-Laverge. La mémoire constructive : repenser la mémoire pour penser l'identité personnelle. Philosophie. Université de Nanterre - Paris X; Universiteit Antwerpen, 2020. Français. NNT : 2020PA100104 . tel-03506264

HAL Id: tel-03506264

<https://theses.hal.science/tel-03506264>

Submitted on 2 Jan 2022

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.

Membre de l'université Paris Lumières
École doctorale 139 : Connaissance, langage, modélisation
Équipe d'Accueil IRePh

Faculty of Arts
Antwerp Doctoral School
Centre for Philosophical Psychology

Lorraine Gérardin-Laverge

Constructive memory

Rethinking memory to redefine personal identity

Thèse présentée et soutenue publiquement le 24/11/2020
en vue de l'obtention du doctorat de Philosophie de l'Université Paris Nanterre
sous la direction de M. Denis FOREST (Université Paris 1 Panthéon Sorbonne) /
M. Bence NANAY (University of Antwerp)

Jury :

Rapporteur :	M. Kourken MICHAELIAN	Professeur, Université Grenoble Alpes
Rapporteuse :	Mme. Sarah K. ROBINS	Associate Professor, University of Kansas
Membre du jury :	Mme. Frédérique DE VIGNEMONT	CNRS Senior Researcher, Institut Jean Nicod.
Membre du jury :	M. Philippe HAMOU	Professeur, Université Paris Nanterre

La mémoire constructive

Repenser la mémoire pour penser l'identité personnelle

Loraine Gérardin-Laverge

Résumé

Dans cette thèse, je propose de repenser la mémoire afin de repenser l'identité personnelle. Je pars de la question suivante : Comment est-il possible que les personnes, malgré les changements qui les affectent, se reconnaissent comme mêmes à différents moments du temps ? Une réponse classique à la question diachronique de l'identité personnelle est qu'elle repose sur la mémoire : les souvenirs fondent notre continuité psychologique. Cependant, les récentes recherches empiriques sur la mémoire épisodique montrent qu'elle a une dimension constructive et qu'elle n'est pas seulement une capacité de stockage fidèle du passé. Qu'est-ce que cela change à la question de l'identité personnelle ? J'explore d'abord la théorie mémorielle de l'identité personnelle de John Locke, et je soutiens qu'être une personne, selon Locke, c'est se reconnaître comme telle à différents moments du temps et donc, dans cet acte d'auto-reconnaissance, constituer son identité personnelle. Je soutiens cependant que la vision conservatrice de la mémoire de Locke doit être révisée et je propose de repenser le concept de mémoire en m'appuyant sur les sciences contemporaines de la mémoire. Je soutiens qu'elle a une dimension constructive et propose une théorie de la mémoire constructive de l'identité personnelle. La mémoire épisodique est à la fois une capacité qui me permet de me reconnaître et, parce que cette reconnaissance n'est pas une simple reconnaissance mais la construction d'une représentation de moi-même à travers la collecte d'informations provenant de diverses sources, elle peut produire et constituer mon identité personnelle.

Mots-clés : Identité personnelle ; Connaissance de soi ; Mémoire épisodique ; Mémoire constructive ; Voyage mental dans le temps

Constructive memory

Rethinking memory to redefine personal identity

Loraine Gérardin-Laverge

Abstract

In this thesis, I propose to rethink memory in order to rethink personal identity. I start from an interrogation about personal identity. How is it possible that people, despite the changes that affect them, recognize themselves as themselves? A common answer to the diachronic question of personal identity is that memory is what makes the self: memories allow us to connect with ourselves and to have an idea of our diachronic personal identity. But interestingly, the recent empirical research on episodic memory shows that it has a constructive dimension and is not only a storage capacity that allows one to preserve and retrieve accurate representations of the past. What does it change for the question of personal identity? I start with an exploration of John Locke's memory theory of personal identity, and argue that to be a person, in Locke's view, is to recognize oneself as same at different moments of time and thus, in this act of self-recognition, to constitute oneself as a person with a temporal dimension. I argue that Locke's preservative view of memory has to be revised, and I propose an empirically informed discussion on the concept of memory. I contend that episodic memory has a constructive dimension and has both epistemic functions and functions related to the constitution of diachronic personal identity. I propose a constructive memory theory of personal identity. Episodic memory is at the same time a capacity which allows me to recognize myself and, because this recognition is not a simple recognition but a construction of a representation of myself through the gathering of information from various sources, it can produce and constitute my personal identity.

Keywords: Personal identity; Self-knowledge; Episodic memory; Constructive memory; Mental time travel

Affiliations

UNIVERSITE PARIS NANTERRE

UFR de Philosophie

Ecole Doctorale Connaissance, Langage, Modélisation (ED 139)

Equipe d'accueil Institut de Recherches Philosophiques (IRePh)

200 avenue de la République

92001 Nanterre Cedex (France)

www.parisnanterre.fr

UNIVERSITY OF ANTWERP

Faculty of Arts

Antwerp Doctoral School

Centre for Philosophical Psychology

Stadscampus - Gebouw S

Lange Sint-Annastraat 7,

2000 Antwerpen (Belgium)

<https://www.uantwerpen.be/en/research-groups/philosophical-psychology/>

Résumé substantiel

Dans cette thèse, j'ai proposé de repenser la mémoire pour repenser l'identité personnelle.

Je suis partie d'une interrogation au sujet de l'identité personnelle. Comment est-il possible que les personnes, malgré les changements qui les affectent, se reconnaissent elles-mêmes comme elles-mêmes ? Comment se fait-il qu'elles ne doutent généralement pas de qui elles sont, au sens qu'elles ne doutent généralement pas d'être continues avec elles-mêmes ? Comment se fait-il que nous vivions nos vies de manière à peu près fluide et que la plupart du temps la continuité du soi, la continuité psychologique, ne soit même pas une question ?

Une réponse répandue à la question de l'identité personnelle consiste à dire que c'est la mémoire qui fait le soi : les souvenirs nous permettent de nous lier à nous-mêmes et d'avoir une idée de nous qui s'étend au-delà de notre flux de conscience présent. C'est donc la mémoire épisodique qui produit la continuité psychologique.

Mais voilà, ce qui m'a intéressée aussi et en particulier au début de cette thèse, c'était les recherches récentes sur la dimension constructive de la mémoire épisodique et sa possible relation à l'imagination du futur en première personne. Dans les années 2007, en neuropsychologie, des chercheur.e.s formulent

l'hypothèse de la simulation épisodique constructive (Schacter et Addis, 2007). Cette hypothèse, que j'ai rencontrée tôt, a motivé cette recherche.

Si la mémoire épisodique est constructive, bien qu'elle ne le paraisse pas, et si elle est essentielle à notre représentation de nous-mêmes et à notre continuité psychologique, qu'est-ce que c'est que se reconnaître comme même ? Est-ce se tromper ? Est-on quelque chose de plus, des personnes au sujet desquelles il s'agirait d'enquêter ? Bref, qu'est-ce que ça change, à la question de l'identité personnelle, que la mémoire ne soit pas ce qu'on a cru qu'elle était ?

Première partie

Je me suis intéressée d'abord à une tradition qui fait de la mémoire l'ingrédient essentiel de l'identité personnelle. Cette tradition débute avec John Locke. Dans la première partie de la thèse, j'ai exploré la théorie lockéenne de l'identité personnelle. Et j'ai d'emblée nourri ma réflexion des recherches empiriques contemporaines sur la mémoire. L'idée de Locke est la suivante : ce qui est essentiel pour se reconnaître comme un soi d'un point de vue synchronique, c'est la conscience, au présent pour s'individuer ; et d'un point de vue diachronique, la conscience d'être la même, encore au présent, pour se reconnaître comme même à différents moments du temps et se donner une épaisseur temporelle. Dans cette seconde dimension, j'ai découvert et défendu, bien qu'elle soit discutée, l'importance de la capacité mémorielle dans la théorie lockéenne. Mais avec elle surgissent de nouveaux problèmes : certains qui ont été soulevés depuis longtemps, comme la circularité et la transitivité de la mémoire, et d'autres plus récents, liés à la dimension constructive de la mémoire épisodique découverte par les sciences contemporaines.

Premier Chapitre

Dans le premier chapitre de la thèse, j'ai cherché à décrire la théorie lockéenne en m'appuyant en particulier sur le chapitre 27, et en discutant ses interprétations. Au

sujet de l'identité personnelle, on pose traditionnellement deux types de questions. Une question épistémique : qu'est-ce qui rend possible la reconnaissance de soi comme soi ? Et une question métaphysique : qu'est-ce qui fait le soi ? Sur quoi repose l'identité personnelle ?

Ces deux questions ont leur source dans la philosophie de John Locke. Et parmi les commentateurs lockéens, il y a un débat pour déterminer à laquelle de ces deux questions il a cherché à répondre. Weinberg, 2012, par exemple, défend une interprétation métaphysique de la théorie Lockéenne de l'identité personnelle. Selon elle, il y a chez Locke un fait métaphysique de la conscience, qui nous est inaccessible et qui pourtant fonde l'identité personnelle. La conscience a deux dimensions. Une subjective, qui me permet de m'identifier à moi-même, et une objective : la même conscience continuée, comme fait objectif, fonde la personne. Newman, 2015, au contraire, a proposé une lecture épistémique de la théorie de Locke, selon laquelle Locke aurait défendu une thèse à propos de la reconnaissance de soi et de l'accès à soi en première personne, sans se prononcer sur des questions métaphysiques.

J'ai cherché à dépasser la dichotomie entre théorie épistémique et théorie métaphysique du soi. Il m'a semblé qu'une lecture fidèle de Locke invitait au contraire à considérer qu'on ne pouvait pas dissocier ces deux questions, et que l'intérêt de l'intuition lockéenne était justement de répondre aux deux questions ensemble. Et plus précisément de répondre à la question métaphysique de l'identité personnelle par celle de la reconnaissance de soi. Être une personne, chez Locke, c'est se reconnaître comme même à différents moments du temps et ainsi, dans cet acte de reconnaissance de soi, se constituer comme une personne ayant une épaisseur temporelle.

Mais la théorie lockéenne de l'identité personnelle a fait l'objet de certaines critiques, dont une objection rendue célèbre par Butler, 1736 : la conscience de soi, au cœur de la reconnaissance de soi, présuppose, et ne peut donc pas constituer, l'identité personnelle. Cette objection a reçu de nombreuses réponses dans la littérature. Dans cette thèse, il s'agissait de voir si la lecture que je proposais était menacée par l'objection de la circularité. Je me suis demandé si la reconnaissance

de l'identité personnelle pouvait avoir lieu en l'absence d'identité personnelle précédant cette reconnaissance. L'alternative que j'ai posée est la suivante : soit la reconnaissance de l'identité personnelle est toujours seulement un acte de reconnaissance de quelque chose qui est déjà, et dans ce cas, se reconnaître présuppose l'identité personnelle, soit la reconnaissance de l'identité personnelle peut arriver sans identité personnelle précédant l'acte de reconnaissance et dans ce cas, la reconnaissance de soi peut fonder et constituer l'identité personnelle. Alors, la reconnaissance de soi ne présuppose pas l'identité de la personne.

J'ai proposé d'envisager le cas que Locke nomme « l'erreur fatale » (fatal error) : un individu se souvient à tort d'avoir fait une action qu'il n'a jamais faite. Mais en s'en souvenant en première personne, il s'identifie avec le sujet de cette action, se reconnaît en elle, et prend la conscience de l'agent de l'action en question pour la même conscience que la sienne maintenant, qui se souvient.

Le passage dans lequel Locke envisage l'erreur fatale fait l'objet d'un vif débat dans les études lockéennes. Pour plusieurs commentateur.rice.s, ce passage rend la théorie lockéenne incohérente. Le fait que je puisse me tromper sur mon identité personnelle implique qu'elle ne peut définitivement pas dépendre du fait que j'y aie accès.

Cependant, j'ai défendu que, chez Locke, l'erreur au sujet de soi-même en tant que personne est une question ambiguë. Si on donne toute son importance à la distinction qu'il opère entre l'identité individuelle, l'identité humaine et l'identité personnelle, elle n'est plus un problème. Dans le cas de l'erreur fatale, je peux me tromper sur mon identification de moi-même en tant qu'individu humain mais je ne peux pas me tromper sur mon identité personnelle, puisque celle-ci dépend de mon accès à elle en première personne. Dans le cas de l'erreur fatale, en fait, je ne fais pas d'erreur du point de vue de mon identité personnelle. En m'identifiant à une personne dans mon souvenir, je constitue mon identité personnelle avec cette personne souvenue.

Locke invoque Dieu dans ce passage et, comme s'il anticipait les critiques, dit qu'au Jugement dernier, Dieu résoudra ces erreurs d'identification en pratiquant ce que

j'ai décrit comme une sorte d'anamnèse platonicienne : Dieu rendra à chacun.e la conscience de ses actions, et ainsi pourront coïncider identité individuelle et identité personnelle. Ce qui est intéressant, c'est que même lorsque Locke invoque Dieu au sujet de l'identité personnelle, il ne lui attribue pas le rôle de décider objectivement des identités des personnes, mais seulement de rendre à chacun.e la conscience de ce qu'il.elle a fait. Autrement dit, même dans ce cas, la réponse à la question de l'identité personnelle doit passer par la conscience du sujet, et la réponse à la question métaphysique dépend de la réponse à la question épistémique. L'identité personnelle dépend de sa reconnaissance par le sujet.

J'ai apporté un argument empirique contre l'objection de la circularité et rapproché le cas de l'erreur fatale de ce qui est nommé aujourd'hui en psychologie les faux souvenirs. J'ai montré que ces cas ne sont pas des cas de science-fiction mais des cas courants où un sujet se souvient avoir fait une action qu'il.elle n'a pas faite. Cependant, dans ces cas, le sujet qui se souvient s'identifie au sujet dont il.elle se souvient. J'ai introduit le concept de conscience auto-noétique, proposé par Tulving, 1985 pour comprendre ces cas : la conscience auto-noétique est le type de conscience qui accompagne le voyage mental dans le temps en première personne. Les souvenirs sont accompagnés par ce type de conscience, selon laquelle le sujet qui se souvient est aussi le sujet de son souvenir. Dans les cas de faux souvenirs, indépendamment de l'exactitude du souvenir, le souvenir est aussi accompagné d'une conscience auto-noétique, c'est même pour cette raison que j'ai l'impression de me souvenir. Mais ces cas sont des cas dans lesquels il n'y a pas d'identité personnelle précédant l'acte de reconnaissance. Puisque les cas d'erreur fatale, ou faux souvenirs, sont possibles, et que dans ces cas, même si je m'identifie à tort au sujet dont je me souviens, je m'identifie quand même, alors se reconnaître comme même et s'étendre par la conscience ne présuppose pas l'identité personnelle, mais peut la fonder. J'ai donc montré que la conscience auto-noétique, c'est-à-dire la conscience de l'identité personnelle dans le temps, puisqu'elle accompagne souvenirs et faux souvenirs, ne présuppose pas l'identité personnelle.

Ainsi, j'ai défendu que la conscience et la mémoire étaient essentielles à la théorie lockéenne de l'identité personnelle, en tant que capacités qui me permettent de me reconnaître et donc de me constituer.

Second Chapitre

Dans le second chapitre de la thèse, j'ai poursuivi cette exploration de la théorie lockéenne de l'identité personnelle en m'interrogeant sur sa conception de la mémoire et le rôle que la mémoire peut jouer dans l'identité personnelle. Ce chapitre s'ouvre ainsi par une étude des passages dans lesquels Locke dessine sa conception de la mémoire et de la capacité à se souvenir (notamment du chapitre II, x de l'*Essai*). J'ai montré que Locke a une conception préservative de la mémoire. Chez lui, la mémoire est supposée préserver les idées qu'on a eues, afin d'être capable de les retrouver plus tard. J'ai tenté de reconstruire une théorie de la mémoire chez Locke et j'ai énuméré les conditions du souvenir. Pour qu'un état mental appartienne à la catégorie souvenir, il doit remplir plusieurs conditions :

- (1) Représentation : Un individu A a maintenant en t1 une représentation d'un événement passé E.
- (2) Conscience antérieure ou *previous awareness condition (PAC)* : A doit avoir eu l'idée de E dans le passé en t-1.
- (3) Conscience autonétique : A est conscient.e que sa représentation de E en t1 est la représentation d'un événement ou d'une idée qu'il.elle a antérieurement perçue en t-1.
- (4) Mêmété : La représentation de E en t1 par A est la représentation de la perception passée et préservée de E en t-1 par A. La représentation de E maintenant et au passé doit être la même idée.
- (5) Véracité (accuracy) : La représentation de E en t1 est une représentation vé-race de ce que A a perçu en t-1

(6) Stockage : E doit avoir été stocké [dans le cerveau de A] quand E a été perçu, sous la forme d'une trace mémorielle.

(7) Causalité : La représentation de E par A en t1 est causée par la perception de E par A en t-1, grâce à la préservation de son impression.

La condition (6) de stockage et la condition causale (7) sont au cœur de la conception lockéenne du souvenir, puisqu'elles sécurisent ensemble la condition de mêmeté et la condition de véracité. Ensemble, ces conditions garantissent que le souvenir est la remémoration de représentations fidèles du passé, et peut donc constituer l'identité personnelle dans la théorie lockéenne.

Mais une seconde objection célèbre, celle de Berkeley, 1732, et Reid, 1785, met cette théorie en péril : fonder l'identité personnelle sur la mémoire mène à un problème théorique de transitivité. La version de Reid, connue comme celle du « brave officier » est la suivante : supposons un brave officier qui a été battu à l'école enfant, pour avoir volé des fruits, qui, soldat, a pris un drapeau à l'ennemi dans sa première campagne et qui a été fait général plus tard. Supposons aussi que quand il a pris le drapeau, il se souvenait avoir volé les fruits. Et supposons que quand il a été fait général, il se souvenait avoir pris le drapeau mais oublié avoir volé les fruits. Selon Locke, il est la même personne aussi loin que sa conscience de soi s'étend. Il s'ensuit que, lorsqu'il était soldat, il était la même personne que l'enfant voleur, et lorsqu'il a été fait général, il était la même personne que le soldat. Cependant, quand il est fait général il ne peut pas être décrit comme étant la même personne que l'enfant voleur puisqu'il ne se souvient pas de cette expérience passée. Ainsi, il est et il n'est pas la même personne que l'enfant voleur. La théorie de Locke souffre donc d'un problème de transitivité.

Je me suis intéressée en particulier à la formulation reidienne de l'objection de la transitivité, car elle est accompagnée chez Reid d'une critique plus générale de la conception lockéenne de la mémoire. Ce n'est pas seulement une critique de sa théorie de l'identité personnelle mais aussi une critique de sa conception préservative de la mémoire. Chez Locke, quand je me souviens, je revis une expérience passée en retrouvant une idée antérieurement formée pendant la

perception et qui a été stockée. Pour Reid, au contraire (et cela s'inscrit chez lui dans un arrière-plan métaphysique de défense d'un réalisme direct), la mémoire ne peut pas être la recollection d'expériences passées mais est une relation directe aux événements expérimentés dans le passé. Il s'oppose à la conception de la mémoire lockéenne comme un entrepôt (store-house), au motif que l'existence d'impressions stockées dans le cerveau manque de preuves. De plus, même si de telles traces étaient possibles, elles pourraient n'être que de simples corrélats et non pas des causes du souvenir. Enfin, puisque Locke dit que les idées cessent d'être quand elles ne sont pas actuellement perçues, la mêmeté entre les idées de perceptions et les idées mémorielles est improbable. Selon Reid, une chose ne peut avoir deux débuts d'existence, et par ailleurs, il accuse Locke de ne pas distinguer entre la perception et la mémoire.

Afin de répondre à ces deux critiques, j'ai d'abord étudié les chemins suivis dans les études lockéennes pour dépasser l'objection du brave officier. Elles sont de deux types. Des philosophes ont tenté d'adapter le concept de mémoire en soutenant que la mémoire liait les idées par une relation d'ancestralité. Le général, en s'identifiant au soldat, est lié aux expériences que le soldat s'est appropriées même s'il ne s'en souvient plus. Une seconde stratégie consiste à diminuer l'importance du rôle de la mémoire dans la théorie lockéenne. Cette voie est empruntée par Atherton, 1983, pour qui ce qui fait l'identité personnelle chez Locke est la conscience, et non la mémoire. Elle est suivie par Shelley Weinberg, 2012, et aussi en un sens par Philippe Hamou, 2014.

Leur lecture a le mérite de régler un grand nombre de problèmes de la théorie lockéenne. Mais j'ai tenté de montrer qu'elle se heurtait à des difficultés pour expliquer les passages où Locke semble considérer la mémoire et la conscience étendue comme une seule et même chose. Et sans mémoire, il n'est pas clair comment un sujet peut accéder au détail de son histoire personnelle. Or, comme je l'ai soutenu, l'accès à soi fonde l'identité personnelle chez Locke.

Enfin, la seconde partie de la critique de Reid est peu considérée dans la littérature, et j'ai montré que les deux voies empruntées dans les études lockéennes au sujet de la mémoire partageaient avec Locke une conception préservative de la mémoire.

Dans les dernières sections du chapitre 2 j'ai proposé une troisième voie pour répondre à l'objection de Reid, en m'appuyant sur des découvertes empiriques récentes. J'ai proposé de repenser le rôle que peut avoir la mémoire dans une théorie lockéenne. Contre Reid, et en m'appuyant sur les sciences de la mémoire, j'ai soutenu qu'il existe des choses telles que des traces mémorielles. Mais elles n'impliquent pas nécessairement la véracité du souvenir, et elles garantissent encore moins la condition de mêmeté. J'ai ainsi proposé une vue hybride de la mémoire épisodique, comme capacité à la fois préservative et constructive.

J'ai montré d'abord que la psychologie de la mémoire depuis la fin du XIXe siècle donne des éléments en faveur de la critique reidienne de la conception préservative de la mémoire. La mémoire ne nous donne pas accès à des images fidèles du passé. Je me suis appuyée sur les recherches sur les témoins oculaires qui ont commencé à la fin du XIXe ([Bartol and Bartol 2006](#)), qui sont poursuivies jusqu'à aujourd'hui ([Loftus 2005](#)) et qui montrent que les témoins oculaires d'une scène ne s'en souviennent jamais parfaitement. Il.elles ajoutent des éléments, en oublient d'autres, et reconstruisent une représentation de l'événement. Assister à un événement n'implique pas qu'on s'en souviennent de manière véridique.

J'ai montré que l'encodage et la récupération étaient tous les deux des processus constructifs. Percevoir une scène n'est pas percevoir la vérité d'une scène, c'est toujours un regard sur une situation qui dépend d'un arrière-plan du sujet, de biais et de l'attention. Pour le montrer, je me suis appuyée sur une expérience de Sophie Calle, sur des études empiriques, et sur la littérature philosophique sur la dimension constructive de l'encodage ([McCarroll, 2018](#)).

J'ai montré aussi que la mémoire n'est pas seulement constructive, mais qu'elle est aussi reconstructive. C'est-à-dire que chaque remémoration est un processus de reconstruction qui donne une nouvelle représentation d'un événement. Les études de Bartlett et sa méthode expérimentale de la Reproduction Répétée en donnent une illustration ([Bartlett, 1932](#)). J'ai soutenu que Reid avait raison de mettre en question la condition de mêmeté, mais que cela n'impliquait pas pour autant d'adopter un

réalisme direct, et exigeait plutôt que nous adoptions une conception constructive de la mémoire épisodique, de la mémoire des souvenirs.

Si j'ai montré que l'encodage et la récupération épisodiques avaient une dimension constructive j'ai aussi défendu que ce n'était pas tout. Et j'ai mis en doute les conclusions de Reid sur la nature de la mémoire. La mémoire ne met pas les gens en relation directe avec les événements passés mais apparaît comme la récupération de traces mnémoriques stockées et encodées. Je me suis appuyée sur la littérature en psychopathologie qui s'intéresse depuis le début du XXe jusqu'à aujourd'hui à l'étude des patient.e.s amnésiques. Un cas célèbre est celui de HM qui, après une lobectomie, sombre dans une amnésie profonde notamment en ce qui concerne ses souvenirs récents. Il a par ailleurs des difficultés à former de nouveaux souvenirs. Le cas de HM n'est pas isolé. Dans des amnésies épisodiques, souvent, une partie du cerveau appelée hippocampe est lésée et la lésion de l'hippocampe résulte toujours dans une forme d'amnésie épisodique. L'hippocampe est nécessaire à la mémoire épisodique. Si, à l'époque de Reid, il n'y a pas de preuves suffisantes pour soutenir que la mémoire a à voir avec le cerveau, la situation est différente aujourd'hui.

J'ai donné des raisons d'adopter une vue hybride de la mémoire épisodique qui la comprend comme à la fois préservative et constructive. J'ai rejeté une conception uniquement préservative de la mémoire, et rejeté la condition de la mêmeté dans une théorie de la mémoire. J'ai donc formulé l'hypothèse que la conception préservative de la mémoire chez Locke était improbable, mais qu'une vue hybride de la mémoire, en revanche, permettait de mieux comprendre le rôle fondamental de la mémoire épisodique dans l'identité personnelle, et d'éviter les objections classiques.

Seconde Partie

Dans la seconde partie de la thèse, j'ai approfondi la réflexion sur le concept de mémoire. Cette partie s'intègre dans une réflexion contemporaine en sciences et en

philosophie de la mémoire sur la capacité mémorielle, et en particulier sur la mémoire épisodique. Dans le chapitre 3 j'ai discuté la nécessité de deux conditions présentes dans la plupart des théories de la mémoire contemporaines : la factivité et la condition causale. Dans le chapitre 4, je me suis intéressée aux fonctions de la mémoire épisodique et j'ai exploré ses fonctions non-épistémiques.

Troisième Chapitre

Le chapitre 3 explore les théories contemporaines de la mémoire. J'y ai notamment discuté la condition de factivité et la théorie causale du souvenir proposée par Martin et Deutscher, 1966, et adoptées par la plupart des théories de la mémoire.

Selon la condition de factivité, un sujet S se souvient d'un événement E si E. Pour qu'un sujet S ait un souvenir M d'un événement E, E doit s'être produit dans le passé. Cette condition garantit que le souvenir est le souvenir d'un événement vrai.

Selon la condition causale, pour qu'un sujet S ait un souvenir M d'une expérience passée Exp, Exp doit avoir été opérative dans la production de M. La condition causale garantit que le souvenir est souvenir d'une expérience authentique.

Dans ce chapitre, j'ai proposé un cas fictionnel, qui d'un côté s'inscrit dans une tradition en philosophie de l'esprit dont Locke est encore un des initiateurs, et dont les néo-lockéens, comme Parfit, 1984, par exemple, sont friands. Mais, d'une part, ce cas ressemble plus à un scénario policier (un meurtre dans une maison de retraite, dont Anna, une infirmière se croit responsable à tort) et, d'autre part, je ne l'ai pas utilisé pour tirer des conséquences théoriques à partir d'intuitions. Je m'en suis servie pour isoler des cas différents dans lesquels la question de savoir si des contenus mentaux sont des souvenirs se pose. J'ai proposé donc de considérer 4 cas :

M1 : Anna se souvient d'avoir donné des somnifères à Ms. F. Cet événement s'est bien produit, et son expérience a pu produire son souvenir. M1 peut remplir les deux conditions.

M2 : Anna se souvient avoir donné de l'hydroxyzine à Ms. F. Or, Anna a bien donné des somnifères à Ms. F mais pas de l'hydroxyzine (la condition de factivité n'est pas remplie). La représentation courante d'Anna n'est pas seulement causée par l'expérience passée mais aussi par des informations qu'elle a reçues a posteriori : elle a trouvé une plaquette d'hydroxyzine vide sur la table de nuit de Ms. F le lendemain (la condition causale n'est pas strictement remplie).

M3 : Anna se souvient d'avoir vu Ms. J dans l'encadrement de sa porte ce lendemain matin quand elle s'est débarrassée de l'emballage d'hydroxyzine vide dans le couloir. Anna a en fait halluciné, Ms. J n'était pas dans l'encadrement de sa porte, Anna a vu une ombre et cru que c'était Ms. J (la condition de factivité n'est pas remplie). La représentation courante d'Anna peut être causée par son expérience passée (la condition causale est remplie).

M4 : Anna se souvient de donner des pilules contre l'allergie à Ms. F. C'est bien ce qu'il s'est passé (la condition de factivité est remplie). Sa représentation courante est le fruit d'une révision mémorielle. Elle est causée par de l'information post-événement et peut-être aussi par la remémoration de l'expérience passée (la condition causale n'est pas strictement remplie).

Si les deux conditions discutées sont nécessaires au souvenir, alors Anna ne se souvient vraiment que dans M1. Dans les autres cas, il s'agit d'autre chose. L'enjeu du chapitre 3 a été de discuter la nécessité de ces deux conditions.

J'ai commencé par discuter la condition de la factivité. J'ai soutenu plusieurs choses. D'une part, elle implique une forme de disjonctivisme qui, en plus de manquer de preuves, est difficile à comprendre dans le cas de M3. M3 est un cas dans lequel Anna se souvient, mal peut-être, mais un cas dans lequel elle se souvient. La condition de factivité ne fait pas de place aux erreurs mémorielles et c'est un problème.

D'autre part, dans les théories du souvenir qui posent la factivité comme une condition nécessaire, elle est souvent corrélée (comme c'est le cas chez Markus et

Werning, 2016) à l'idée que les faux souvenirs ne sont en fait pas des souvenirs. Ils ne sont pas des souvenirs car ils ne respectent pas la condition de factivité. J'ai montré que cette idée reposait sur un malentendu. Les recherches sur les faux-souvenir, en fait, ne permettent pas de décider si les faux souvenirs sont vrais ou faux par rapport à l'évènement passé qu'ils représentent. Elles sont plutôt reliées à la condition causale de la factivité, puisque les faux souvenirs sont le résultat de l'intégration d'information post-événement dans la représentation de l'évènement. L'information en question peut être trompeuse (comme c'est le cas dans M2), mais elle peut aussi être vraie (comme c'est le cas dans M4) et permettre une révision d'un souvenir distordu. Si les conditions de factivité et causale sont souvent présentées comme interdépendantes, j'ai montré qu'elles ne sont pas nécessairement remplies ensemble. Le problème des faux souvenirs est davantage relié à la condition causale qu'à la condition de factivité. Les souvenirs issus de la reconstruction d'évènements passés avec une intégration d'information post événement ne sont pas toujours faux.

Enfin, j'ai soutenu que poser la condition de factivité comme condition nécessaire au souvenir présupposait en fait une conception biaisée de la mémoire épisodique. La condition de factivité n'est pas nécessaire pour se souvenir, elle est nécessaire pour connaître le passé sur la base de la mémoire. C'est-à-dire que considérer la condition de factivité comme une condition nécessaire au souvenir implique qu'on a déjà décidé à quoi servait la mémoire épisodique : à connaître le passé. J'ai soutenu que ce présupposé devait être questionné (et c'est l'objet du chapitre 4).

Dans la troisième section du chapitre 3, j'ai discuté la condition causale. La théorie causale du souvenir est endossée par la plupart des théories contemporaines du souvenir, mais elle a récemment fait l'objet d'un débat. Elle a été posée initialement par Martin et Deutscher (1966). Et je l'ai expliquée comme suit :

Condition causale : Le souvenir M de S de l'expérience passée Exp doit être causé de manière appropriée i.e. par une chaîne causale non déviante, par Exp.

Selon la condition causale, M2 et M4 ne sont pas des souvenirs. Qu'ils soient vrais ou faux, ils ne sont pas causés de manière appropriée par l'expérience passée, mais ont une autre source. M1 peut être défini comme un cas de souvenir, car il peut être

causé par l'expérience passée. M3 aussi. Bien qu'il soit faux par rapport à l'événement passé, il peut être authentique par rapport à l'expérience passée. Anna a halluciné mais elle a bien eu cette expérience. Son souvenir d'hallucination pourrait être causé par son expérience passée.

J'ai montré que dans M4, selon la théorie causale, Anna croit qu'elle se souvient mais elle commet une erreur de « source-monitoring » (Johnson et al. 1993). C'est-à-dire qu'elle forme un jugement métacognitif erroné sur la source de sa représentation du passé : elle croit que cette dernière vient de son expérience passée, alors qu'elle vient en fait d'information post-événement. M2 est le résultat d'un effet de désinformation. Sa source n'est pas l'expérience passée représentée, mais il est causé par de l'information post-événement. M4 est vrai (par rapport à l'événement passé) mais inauthentique (par rapport à l'expérience passée), M2 est faux (par rapport à l'événement passé) et inauthentique (par rapport à l'expérience passée).

Dans la section consacrée à la théorie causale, j'ai soutenu que la binarité entre faux-souvenir et souvenir authentique n'aide pas à expliquer la mémoire épisodique.

J'ai d'abord montré qu'il était courant de se souvenir de davantage d'éléments que de ce qui est présent dans une expérience passée représentée, en m'appuyant notamment sur le paradigme DRM (Roediger et McDermott, 1995). Le fonctionnement normal de la mémoire épisodique implique une activité de remplissage de trous, de recomposition. L'idée que l'expérience passée doit être la cause du souvenir n'est pas vraiment compatible avec cette compréhension de la mémoire épisodique.

J'ai montré que l'encodage et la récupération sont tous deux des processus qui s'appuient sur des informations provenant de diverses sources. D'une part, l'encodage n'est pas causé seulement par l'expérience passée, puisque la perception implique des processus de complétion, comme c'est le cas dans le phénomène robuste de l'extension des limites (boundary extension). D'autre part, la récupération en mémoire épisodique s'appuie aussi sur la collecte d'informations

de sources diverses. En particulier elle s'appuie aussi sur des informations sémantiques, sur ce que le sujet qui se souvient sait par ailleurs.

J'ai soutenu que la mémoire épisodique interagit avec d'autres types d'information, de sorte qu'il est difficile de dire qu'un souvenir épisodique retrouvé vient seulement de l'expérience passée qu'il représente. Soit on a une version faible de la condition causale, selon laquelle le souvenir de Exp doit être causé partiellement par Exp, mais dans ce cas, elle n'est pas très explicative. En effet selon cette version faible, M1, M2, M3, M4 peuvent compter comme des souvenirs, ils peuvent tous être décrits comme partiellement causés par l'expérience passée qu'ils représentent. En un sens, une telle condition perdrait son pouvoir explicatif. Soit on a une version stricte de la condition causale, et dans ce cas, en fait, aucun des cas M1, M2, M3, M4 ne peut vraiment compter comme souvenir. Ils peuvent tous être décrits comme résultant de la collecte d'informations de sources diverses. Une telle condition perdrait aussi son pouvoir explicatif.

J'ai donc défendu que la condition de factivité et la condition causale (dans sa version stricte) ne pouvaient pas être considérées comme des conditions nécessaires pour se souvenir.

Quatrième Chapitre

Ces deux conditions jouent un rôle important dans les théories épistémiques de la mémoire épisodique car elles sont censées garantir, pour l'une, la vérité du souvenir par rapport à l'événement passé qu'il représente (factivité), et pour l'autre, l'authenticité du souvenir par rapport à l'expérience passée qu'il représente (condition causale). Mais voilà, elles présupposent que la fonction principale de la mémoire épisodique est de représenter le passé de manière vraie et authentique.

Dans le chapitre 4, c'est ce présupposé que j'ai discuté. La question des fonctions de la mémoire épisodique fait l'objet d'un vif débat dans les recherches récentes sur la mémoire en sciences cognitives et en philosophie. Des chercheur.e.s ont récemment défendu que la découverte d'une dimension constructive de la mémoire,

en s'intéressant aux erreurs mémorielles, conduisait à oublier ce à quoi la mémoire épisodique sert : connaître le passé. Selon eux, une question centrale sur la mémoire tient à son évaluation, et on évalue les souvenirs en fonction de leur valeur épistémique sur le passé.

J'ai commencé la discussion en explicitant leurs arguments. J'ai montré que cette thèse avait été défendue de trois façons dans la littérature récente. Chez Mahr et Csibra, 2018, il s'agit de la condition causale. Selon eux, la recherche sur la dimension constructive de la mémoire épisodique est reliée à un projet descriptif à l'intérieur de la psychologie, qui décrit ce que la mémoire fait. Mais cela ne devrait pas rentrer en compétition avec ce que la mémoire est censée faire. Pour deux raisons : d'une part, les souvenirs se présentent au sujet comme étant causés par leur expérience passée. Le souvenir (output) semble être causé par l'expérience passée (input). D'autre part, il y a deux façons différentes de se poser la question de la nature de la mémoire. L'une est descriptive, elle se concentre sur la meilleure description de la mémoire épisodique. L'autre est normative et se concentre sur les conditions qui justifient qu'on considère un état mental comme un souvenir. Si on comprend que la théorie causale est engagée dans un projet normatif et pas descriptif, alors les arguments empiriques contre la théorie causale ne tiennent pas.

Une seconde façon de soutenir cette thèse est proposée par Henry et Craver, 2018 et Craver, 2020 : la *mémoire normative* a des conditions de succès qui ne dépendent pas de la mémoire empirique. La mémoire normative a des conditions de succès qui sont épistémiques. Et la règle numéro un c'est qu'un souvenir correspond à une expérience de première main. La personne qui se souvient a une autorité épistémique particulière. Si la recherche empirique montre que les souvenirs peuvent avoir d'autres sources que l'expérience de première main, ça ne change pas le fait que les évaluations mémorielles en dépendent. L'expérience de première main a une valeur épistémique importante.

Une troisième voie a été proposée par Barner, draft, et se concentre sur le discontinuisme normatif. Bien que la recherche empirique sur le voyage mental dans le temps mène certains à adopter une thèse continuiste (Michaelian, 2016a) et à considérer le voyage mental dans le temps orienté vers le passé et le voyage

mental dans le temps orienté vers le futur comme étant de la même sorte (kind), il y a une discontinuité fondamentale d'un point de vue normatif entre se souvenir du passé et imaginer le futur. Les souvenirs et les simulations du futur ne peuvent pas être évalués épistémiquement de la même manière. Les souvenirs sont factifs, ils ont eu lieu, ils ont des conditions de vérité. Les simulations du futur, quant à elles, n'ont pas eu lieu, elles concernent des choses qui ne se sont pas encore produites. On évalue les souvenirs en fonction de leur véracité par rapport au passé, mais on ne peut pas évaluer les simulations du futur de la même manière.

Contre ces trois versions, j'ai essayé de défendre que la mémoire épisodique n'a pas seulement des fonctions épistémiques. Ces arguments sur la mémoire normative et la mémoire empirique cachent le fait que la mémoire épisodique n'est pas tout entière tournée vers la connaissance du passé vrai.

Mon premier argument contre ce que j'ai appelé la thèse épistémico-normative, c'est que les normes d'évaluation de la mémoire ne sont pas uniques. J'ai exploré les cas de distorsions cliniques de la mémoire. J'ai soutenu que dans les cas où les distorsions mémorielles sont épistémiquement innocentes (Bortolotti, 2015), c'est-à-dire où la personne qui se souvient ou confabule ne pourrait pas former une représentation plus authentique du passé, elles peuvent avoir des bénéfices épistémiques. Dans ces cas, les normes de leur évaluation ne sont pas les mêmes que dans les cas dits normaux. D'autres règles s'appliquent qui ne sont pas celles de la factivité ou de la causalité. Je me suis appuyée sur la littérature qui montre que les confabulations cliniques ont des bénéfices au moins en deux sens : elles préviennent l'absence nocive de représentation de soi et de confiance en ses capacités, et elles sont une manière d'entraîner des compétences épistémiques (construire une représentation du passé, la formuler, la confronter au jugement d'autrui pour éventuellement la réviser).

Mon second argument a été de montrer plus généralement que les sciences de la mémoire, en découvrant sa dimension constructive, invitaient à réviser les manières de l'évaluer épistémiquement. J'ai défendu que, puisque la construction mémorielle dépendait des contextes de l'encodage et de la récupération, si on s'inscrit dans un projet épistémique de connaissance du passé par la mémoire, la mémoire

personnelle n'est pas suffisante. J'ai soutenu que le domaine empirique peut guider le domaine normatif. Si on veut connaître le passé sur la base de reports mémoriels, cela doit prendre la forme d'une enquête.

Mon dernier argument a été de défendre que le régime épistémique n'est pas le seul régime normatif qui guide la mémoire et les évaluations mémorielles. J'ai montré que on n'évalue pas dans tous les contextes la mémoire épisodique en fonction de sa capacité à représenter fidèlement le passé. Il y a, au contraire, des contextes dans lesquels la fonction épistémique de la mémoire épisodique n'est pas ce qui importe. Si c'est le cas dans un contexte légal par exemple, ce n'est pas le cas tout le temps. Dans la dernière section du chapitre 4, j'ai exploré des évaluations de la mémoire épisodique non-épistémiques : les évaluations mémorielles relatives aux bénéfices psychologiques et aux potentialités d'action. J'ai défendu que la mémoire épisodique avait des fonctions non-épistémiques.

Je me suis appuyée à nouveau sur la recherche sur les confabulations, et j'ai montré que les souvenirs, indépendamment de leur véracité ou de leur authenticité, c'est-à-dire aussi indépendamment de leur coût épistémique, pouvaient avoir des bénéfices psychologiques. Les tentatives de contester les confabulations cliniques sont presque toujours sans succès, et non bénéficiels (Fotopoulou, 2008). Malgré leur coût épistémique, des chercheur.e.s ont montré que les confabulations cliniques avaient des bénéfices psychologiques. Confabuler permettrait notamment d'améliorer la confiance en soi, un sens de la compétence, de la cohérence et un bien-être psychologique (Bortolotti et Sullivan Bisset, 2018).

J'ai par ailleurs proposé que certains souvenirs ne sont pas à évaluer en fonction de leur fidélité au passé mais en fonction de leur effet par rapport à des potentialités d'action. J'ai proposé deux souvenirs d'enfance pour illustrer cet argument.

Souvenir 1 :

Une fête d'école. On joue autour d'une rampe de skate avec des ami.e.s. Le jeu est de sauter dans l'herbe du haut de la rampe. C'est haut. Les gens disent à ma sœur, « Saute Mona, saute ! », elle hésite. Et là, je ne sais pas si c'est par fierté ou sacrifice

mais j'y vais d'un coup, je fais le saut, et je m'étale au sol, j'ai mal au dos et partout et je pleure. C'est un échec.

Souvenir 2 :

Les vacances d'été. On passe beaucoup de temps à la piscine du village avec des ami.e.s. Un jour, on joue à qui ira le plus loin en apnée. Et je fais un aller-retour complet. C'est un succès.

L'un est un souvenir d'échec, l'autre un souvenir de réussite. J'ai montré que l'un est à la base de mon sens de la compétence dans certaines tâches, et l'autre est à la base de mon sens du danger dans certaines situations. Ce qui compte pour moi dans ces souvenirs, ce n'est pas leur véracité mais ce qu'ils me permettent de faire, et comment ils constituent pour moi une base de mon identité personnelle et de mon sens de la compétence.

En ce sens, les arguments que j'ai proposés dans ce chapitre ne sont pas dirigés seulement contre la thèse épistémico-normative, mais aussi plus généralement contre les théories épistémiques de la mémoire épisodique qui sont en fait dominantes dans la philosophie de la mémoire contemporaine. Même chez Michaelian (2016c), qui offre une des critiques les plus sévères de la théorie causale de la mémoire, la mémoire épisodique est tournée vers la connaissance fiable du passé. J'ai tenté de montrer que ce que ces positions ont en commun, c'est de présupposer que la fonction principale de la mémoire épisodique est la connaissance du passé. J'ai défendu, au contraire, que ce n'était pas sa seule fonction.

Troisième Partie

Dans la dernière partie de la thèse, je suis revenue à la question de l'identité personnelle. J'ai tenté de faire le point sur ce que pourrait être une théorie mémorielle de l'identité personnelle si on comprend la dimension constructive de la mémoire épisodique. En effet, l'image de la mémoire épisodique qui émerge de la recherche empirique est l'image d'une capacité constructive, dépendante en

particulier des contextes dans lesquels elle se construit. J'ai proposé de retourner à la théorie lockéenne explorée dans la première partie de la thèse et fait un point sur les pièces manquantes de sa théorie de la mémoire à la lumière de ce que la mémoire épisodique est en fait.

Cinquième Chapitre

J'ai montré d'abord qu'il manquait à Locke de prendre en compte la dimension constructive de la mémoire épisodique. Locke a une vision préservative de la mémoire, et c'est un problème à la fois pour sa théorie de la mémoire et pour sa théorie de l'identité personnelle.

J'ai défendu ensuite que la théorie lockéenne réduisait le soi à son individualité. Elle pose une personne indépendante d'un monde et d'un contexte, une personne isolée seule avec sa conscience. Cette critique, déjà formulée par Leibniz (*Nouveaux Essais*, II ; xxvii), me semble confirmée par ce qu'on sait aujourd'hui de la mémoire personnelle. Les souvenirs sont formés et retrouvés dans un monde social. Non seulement ils sont sensibles aux contextes dans lesquels ils se forment et se récupèrent, mais en plus, les personnes qui se souviennent sont rarement seules quand elles se souviennent, et étaient rarement seules dans leurs expériences. Dans ces cas, un même événement peut être retrouvé par plus d'une personne. Et quand on se souvient collectivement du passé, communément, on ajuste nos souvenirs aux versions des autres. Il y a une dimension sociale du souvenir qui est manquante chez Locke et pourtant essentielle aux processus mémoriels et à l'identité personnelle.

Enfin, j'ai soutenu que la question de l'identité personnelle n'est pas seulement reliée à l'extension de soi dans le passé, mais aussi à sa survie dans le futur. En m'appuyant sur la littérature en psychopathologie qui explore les conséquences de l'amnésie épisodique, j'ai montré que la continuité psychologique ne dépend pas seulement de la remémoration du passé en première personne, mais aussi de la projection dans le futur en première personne. Un problème pour des patient.e.s atteint.e.s d'amnésie épisodique, c'est aussi qu'ils.elles ont des difficultés à se projeter dans le futur en première personne, comme l'a montré Tulving, 1985. Cette

dimension d'appréhension du futur, l'importance de la relation à soi dans un temps subjectif étendu allant du passé au futur de la personne, est encore un maillon manquant de la théorie lockéenne de l'identité personnelle.

Dans la section 2 j'ai exploré la théorie narrative de l'identité personnelle proposée par MacIntyre, 1981, pour les raisons suivantes : d'une part, les pièces manquantes de la théorie lockéenne sont présentes chez MacIntyre. D'autre part, si l'identité personnelle dépend de la mémoire et que la mémoire est une capacité constructive, dépendant des projections dans le futur, et avec une importante dimension sociale, elle pourrait être plus proche que prévu d'une capacité narrative. Je me suis concentrée sur la théorie narrative de MacIntyre davantage que sur des versions plus récentes de la théorie narrative. À ce titre, il est possible que certaines objections que je formule dans le chapitre, ne soient pas valables pour toutes les théories narratives. Mais je me suis concentrée sur MacIntyre pour au moins deux raisons : d'abord parce qu'il est l'un des premiers à proposer une telle vue de l'identité personnelle. Ensuite parce que certains concepts qu'il a formulés, comme celui d'espace partagé des récits, d'enchâssement des récits, et d'ajustement narratif, m'intéressaient particulièrement. Je considère qu'ils peuvent être utiles à une meilleure compréhension de la construction en mémoire épisodique, ainsi que de la constitution et de la reconnaissance de l'identité personnelle.

Chez MacIntyre, l'identité des personnes se fonde sur l'appartenance des épisodes de leur vie à un tout plus large dont ils.elles peuvent rendre raison et dans lequel ils.elles sont inscrites. L'unité d'une vie dépend de l'intelligibilité des actions d'une personne, dans leur mise en cohérence à l'intérieur d'un espace narratif. Mais les récits ne sont jamais fixes. Il y a toujours une multitude de descriptions possibles d'une séquence d'actions, et différentes versions peuvent en être données à différents moments du temps en fonction de ce qui est à expliquer plus spécifiquement. Une capacité narrative est donc une capacité constructive et reconstructive.

Il y a ainsi une multitude de récits de soi possibles, et le meilleur récit est construit dans un dialogue. Par ailleurs il est toujours enchâssé dans des cadres selon lesquels il prend forme. J'ai donc montré que la théorie narrative répond à l'insuffisance

individualiste de la théorie lockéenne. Mon récit de moi-même est enchâssé dans les récits de mes proches et de mon temps. Ces cadres eux-mêmes ont leur histoire propre. La manière dont je construis mon récit de moi-même dépend des contextes dans lesquels il se forme.

Au-delà des contextes dans lesquels mon récit de moi-même est enchâssé, je suis aussi inclus.e dans un monde social. Les autres peuvent ajouter des choses à mon récit, comme l'a suggéré Leibniz (*Nouveaux Essais*, II, xxvii), ils.elles peuvent aussi produire un récit de moi-même différent du mien, et ainsi m'amener à modifier mon récit de moi-même. Quand je produis un récit de moi, il entre dans l'espace partagé des récits dans lequel les gens peuvent donner et demander des raisons pour ce qu'ils.elles ont fait. Un récit de soi est toujours l'objet de révisions et d'ajustement. La construction de soi-même dans le récit de soi n'est donc pas une activité individuelle, mais elle est enchâssée dans un monde social.

J'ai montré qu'un dernier ingrédient essentiel à la théorie narrative de MacIntyre est le concept d'intention. Selon MacIntyre, pour que des épisodes soient des épisodes de ma vie et de mon histoire, ils doivent faire sens dans la continuité de mes intentions. Ces intentions peuvent changer, mais pour que les épisodes passés soient intelligibles pour moi et les autres et puissent fonder mon identité personnelle, ils doivent pouvoir être explicables dans un récit de moi, et dans le contexte dans lequel ils ont eu lieu. Avec le concept d'intention, la théorie de MacIntyre permet de comprendre la continuité psychologique plus largement que par rapport au seul passé.

Si la théorie de MacIntyre semble pallier les insuffisances de la théorie lockéenne, j'ai soutenu qu'elles avaient des présupposés communs. Le concept de récit se pose comme une alternative radicale à celui de mémoire, et pourtant il m'a semblé que c'était cette alternative qui devait être dépassée. Celle qui oppose la vérité de la mémoire au constructivisme du récit, continue de présupposer que la mémoire est une capacité d'enregistrement fidèle du passé. Et pourtant, comme je l'ai montré dans les chapitres précédents, l'image de la mémoire épisodique qui émerge de la recherche contemporaine est celle d'une capacité constructive.

Dans la section 4 du chapitre, j'ai donné des raisons de ne pas adopter une théorie narrative de l'identité personnelle. J'ai rappelé les arguments de Galen Strawson et défendu que le fait que des individus humains n'aient pas de rapport narratif à eux-mêmes est suffisant pour réfuter une théorie narrative de la personne.

J'ai proposé deux cas supplémentaires néanmoins, issus de la littérature empirique, que j'ai posés comme arguments contre la théorie narrative de la personne. Celui des souvenirs sans croyance et le cas de B, patient amnésique décrit par Störing.

J'ai d'abord attaqué l'idée de la cohérence de la vie mentale en m'appuyant sur la littérature sur la croyance. J'ai montré que les gens ont des croyances contradictoires et même peuvent avoir des souvenirs sans croyance (nonbelieved memories), c'est-à-dire des souvenirs épisodiques qui ne sont pas accompagnés par la croyance que ce dont je me souviens m'est arrivé. Le cas des souvenirs sans croyance est difficile à comprendre, puisque le souvenir semble s'accompagner nécessairement dans sa phénoménologie de la croyance qu'il est vrai. J'ai cependant montré que si on abandonne le mythe de la cohérence de la vie mentale, on a de meilleures chances de faire sens d'un tel phénomène.

Le second cas que j'ai proposé est celui de B, patient de Störing, et discuté par Craver, 2012. Craver présente le cas de B comme un cas de personne ayant une amnésie épisodique. Il utilise ce cas contre les théories mémorielles de la personne et pour montrer que des individus atteints d'amnésie épisodique peuvent maintenir un sens de soi et une continuité psychologique. Mais le cas qu'il décrit me semble en fait être un cas de mémoire épisodique. Il me semble que c'est un cas dans lequel B a un souvenir épisodique ayant une expression comportementale et pas déclarative. Le cas est le suivant : B entre dans une salle d'hôpital dans laquelle il a eu, par le passé, une expérience traumatisante. Et bien qu'il n'ait pas de souvenir de cette expérience, relevant de la mémoire déclarative, il s'enfuit. J'ai montré que dans ce cas, B a en fait une réponse comportementale qui est liée au souvenir d'un épisode passé précis, comme en mémoire épisodique. Ce n'est pas une réaction conditionnée qu'il a apprise, elle est liée à un événement particulier. En ce sens, j'ai montré qu'un individu pouvait maintenir une forme de continuité à soi sans passer

par la mémoire explicite et sans passer par le récit, mais que le lien mémoriel était important.

Dans la dernière section du chapitre, j'ai tenté de proposer une théorie de l'identité personnelle qui serait fondée sur la mémoire constructive. J'ai reposé les questions épistémique et métaphysique de l'identité personnelle. J'ai montré que si la mémoire est constructive elle peut jouer le rôle constitutif que Locke lui prête. En étant une capacité flexible de reconstruction et de recombinaison, en fonction de traces d'expériences préservées et de leur recombinaison en fonction des contextes de la situation présente (ce que je cherche à expliquer, ce que je cherche à faire). Se reconnaître par la mémoire est ainsi se constituer. J'ai donc montré qu'un concept de mémoire reconstructive était en fait plus à même de jouer un rôle fondateur dans une théorie lockéenne de l'identité personnelle. La mémoire est constitutive d'un point de vue métaphysique : être une personne c'est se reconnaître soi-même et construire une représentation de soi grâce à la mémoire. Elle est aussi constitutive dans le détail : se reconnaître par la mémoire c'est se constituer dans le détail de son identité personnelle : ce qu'on a fait, qui on est, ce qu'on peut faire et quelles sont nos intentions.

J'ai montré qu'une telle théorie ne tombait plus sous le coup de l'objection de la circularité. J'ai montré que les faux-souvenirs pouvaient aussi jouer un rôle constitutif dans l'identité personnelle. J'ai soutenu que même les confabulations cliniques pouvaient jouer ce rôle. J'ai défendu que, dans le contexte d'une réflexion sur l'identité personnelle, on avait de bonnes raisons de réduire la distinction entre se souvenir et confabuler. Je me suis appuyée encore sur la littérature sur les confabulations. Mettre en doute les confabulations cliniques est souvent nocif. La résistance au doute est même souvent considérée comme une caractéristique de la confabulation clinique. Mais les mettre en doute, en plus d'être inutile puisque les sujets révisent rarement leurs représentations du passé, n'est pas bénéfique. Il a été montré que cela provoque plus souvent une détresse chez les sujets et que ça empire la méfiance qu'ils.elles ont dans les autres. Dans les cas où les sujets ne peuvent pas ajuster leurs représentations à celles des autres, les mettre en doute est néfaste. Cela semble montrer que les confabulations jouent un rôle dans la représentation de

soi et la constitution de soi-même, et que se reconnaître et être reconnu est essentiel à la continuité psychologique.

J'ai enfin discuté la question des sois amnésiques. J'ai défendu qu'une version révisée de la théorie mémorielle, qui prend en compte la dimension constructive de la mémoire, changeait quelque chose au problème. Les théories mémorielles avaient à soutenir que les personnes amnésiques ne peuvent pas maintenir une continuité psychologique, ce qui est une thèse difficile à soutenir. Craver, 2012, a au contraire montré que des personnes peuvent maintenir une continuité psychologique malgré une amnésie épisodique.

J'ai soutenu plusieurs choses en réponse. D'une part, il faut distinguer entre les cas d'amnésie avec et sans confabulation. La confabulation peut jouer un rôle mémoriel et permettre aux personnes de s'étendre, et il n'est pas évident que, dans les cas d'amnésie épisodique, les personnes n'auraient pas une capacité de reconnaissance d'elles-mêmes qui reposerait sur des informations stockées et recombinaison. Ensuite, puisque les personnes ne sont pas isolées, mais comprises dans un contexte social peuplé de personnes en qui ils.elles ont plus ou moins confiance, il est possible que ces dernières puissent suppléer aux mémoires manquantes. Enfin, puisque j'ai montré que la reconnaissance de soi est cruciale indépendamment de la fidélité au passé pour le bien-être psychologique, j'ai défendu qu'il pourrait être utile de changer nos pratiques vis-à-vis des personnes qui confabulent. Plutôt que de se concentrer sur la fidélité au passé et de rejeter les confabulations pour ce qu'elles ne peuvent pas faire, il pourrait être utile de se concentrer sur ce que les confabulations permettent de faire, c'est-à-dire de construire un sens de la continuité psychologique.

Conclusion

J'ai donc défendu dans cette thèse que l'identité personnelle ne dépend pas de la vérité de soi, mais qu'elle repose sur une relation de reconnaissance de soi, de son passé et de ses intentions, qui constitue une représentation de soi pour soi. La

mémoire constructive peut jouer ce rôle double : épistémique et métaphysique. Elle est à la fois une capacité qui me permet de me reconnaître et, parce que cette reconnaissance n'est pas une simple reconnaissance d'un objet qui la précède, elle peut me constituer, en tant que personnes avec une histoire, des capacités, des intentions. J'ai donc en fait montré plusieurs choses : d'une part, que le concept de mémoire devait être repensé dans une théorie mémorielle de l'identité personnelle. D'autre part, que le concept de soi devait lui aussi être repensé. La personne n'est pas un objet dont le contenu serait fixé indépendamment de son appréhension de soi. Cette appréhension produit quelque chose, elle produit son objet. L'identité personnelle est le résultat d'une reconnaissance de soi constitutive. J'ai soutenu que la question de la vérité de la mémoire comme de la vérité du soi étaient à dépasser. Elles peuvent en fait être secondaires dans une réflexion sur l'identité personnelle. Si les questions de vérité, de fidélité au passé et de succès mémoriel sont au centre des discussions sur la mémoire, et sont présentées dans la littérature comme des caractéristiques essentielles de la bonne mémoire, j'ai montré qu'elles n'étaient pas essentielles dans tous les contextes de réflexion sur la mémoire. La mémoire sert aussi au soi. Et quand elle sert au soi, ce n'est pas seulement pour ses capacités préservatives, mais aussi pour ce qu'elle rend possible en termes de recombinaison et de reconstruction. C'est son caractère dynamique qui permet aux individus de se constituer comme des personnes.

En ce sens, le spectre des personnes s'étend par rapport aux théories mémorielles classiques. La théorie mémorielle que j'ai proposée permet de penser que des personnes qui se souviennent mal, voire qui confabulent, sont aussi des personnes parce qu'elles ont cette capacité à se reconnaître et ainsi à se constituer.

Acknowledgments

All my gratitude goes first to my supervisors. I had the great chance to meet Denis Forest when I arrived at University Lyon 3, after a khâgne. He introduced me to the philosophy of mind, philosophy of science, and philosophy of psychology. It is in his class that I discovered neuroscience, and empirically informed philosophy. He was passionate and transmitted this passion to me. It is in his class that I discovered the fascinating works by Schacter and Addis, 2007, at the basis of my desire to do a PhD. For all he gave me, for our collaborations, for fascinating discussions, for his support these last ten years, I thank him sincerely.

Later in my PhD, I had another incredible chance, to meet Bence Nanay. He taught me a lot, about philosophy of mind, about philosophy of psychology, about the academic field. And my thesis grew up with him. For trusting me as a philosopher, for giving me confidence, for the wonderful feedback, for welcoming me in the Centre for Philosophical Psychology, I thank him sincerely.

My gratitude also goes to my jury, who accepted this thesis for examination. I am grateful and proud. Each of them has played a crucial role in my PhD studentship. For welcoming me in the Centre for Philosophy of Memory, making me part of this field, giving me the chance to teach a class on my research, for being intellectually challenging though incredibly generous and supportive, for agreeing to be part of the jury and to discuss my work, I thank Kourken Michaelian sincerely.

I met Frédérique de Vignemont very early in this PhD. She gave me great advice about philosophy and academia, for welcoming me in her office at Jean Nicod, for

taking time to discuss with me, for agreeing to be part of the jury, and to discuss my work, I thank her sincerely.

I read Sarah Robins before meeting her. For her fascinating, challenging and inspiring works, for interesting discussions, and for agreeing to be part of the jury and to discuss my work, I thank her sincerely.

I met Philippe Hamou in the University of Paris Nanterre, and I had the chance to discuss Locke's ideas with him. For his generosity, his welcoming, for fruitful discussion, for encouraging me to develop my own reading of Locke, for agreeing to be part of the jury and to discuss my thesis, I thank him sincerely.

All my gratitude goes to the institutions and staff that supported my research. The IRePh, the ED 139, the University of Paris Nanterre, and in particular, Christine Marin, Catherine Tesnieres, Sylviane Peniarbelle, Thierry Meyer; the doctoral school of Antwerp, the University of Antwerp and the Centre for Philosophical Psychology, the BOF, Simone Krammer and Hanne de Roeck.

I thank Denis Perrin who has welcomed me in Grenoble years ago, encouraged my research, for his energy, kindness, his smart and sharp feedback, for his generosity, I am more than grateful.

I thank Jean-Michel Salanskis who was in my master's jury and who has been a great advisor.

I thank Carl Craver, for his encouragements very early on in my PhD. I presented my project in front of him, the first year of my PhD in Bochum in a Summer School, he was supportive then and has remained as supportive ever since.

I thank Claire Etchegaray for her support and the great feedback she gave me in many occasions.

I thank Francis Eustache and Béatrice Desgranges, for their incredible support very early on, and for agreeing to discuss their neuropsychological point of view with an apprentice philosopher.

I also want to thank Markus Werning and Kevin Reuter for organizing the ECE Summer School 'Memory and mind' in Bochum, where I learnt a lot, and had the chance to attend, among others, Carl Craver's, Lynn Nadel's, Fabrice Teroni's, classes. There, I also had the chance to meet Chris McCarroll, Mostafa Nashaat, Anne-Lise Saive, Marina Trakas and other smart and kind PhD students. This was one of the best experiences during my PhD.

I am grateful to Mona Gérardin-Laverge for incredible discussions and unconditional support, for collaboration in the organization of the International conference Re-member, and on the paper we co-authored and published in *La Revue de Neuropsychologie*.

My gratitude also goes to the International Society for MacIntyrean Enquiry, and especially to Jeffery Nicholas.

This thesis and my PhD years would have been very different without the group Hexis, in Lyon, and in particular Samuel Lepine who has been an incredible advisor and friend. To Marion Lebidan, for having created this seminar, with Samuel Lepine and others, and welcomed me in, to Sabine Collardey who has also been an incredible advisor and friend, to them and to the others with whom we organized the seminar for a couple of years, and to the philosophers who accepted our invitations and gave among the best philosophical presentations I have attended in Lyon, I am very grateful.

I am grateful to the group NormaStim. And in particular to Marie Gaille and Baptiste Moutaud, for welcoming me in the group and encouraging my research.

I am grateful to the research group Mind, at the University of Paris Nanterre and in particular, to Alexandre Charrier for great collaboration, Hortense de Villaine, Anthony Ferreira.

I am more than grateful to the members of the Centre for Philosophical Psychology. I learnt a lot and I received unvaluable feedback. Among them, I am especially grateful to Gerardo Viera who has been one of the best advisors and a great friend

these past two years, to Chris McCarroll, who has been here for advice, support, bibliographical help, and who has been a great friend too. Huge thanks to the members of the 2018-2019 group and among them to Alma Barner, Magdalini Koukou, Constant Bonard, Kevin Lande, Dan Williams, Nick Wiltsher, for their incredible feedback and friendship. Loads of thanks also to the members of the 2019-2020 group, and in particular to Brad Saad, Adam Bradley, Nic Porot, Laura Silva, Allert Von Westen, Kris Goffin, Stephen Muller, Nicolas Alzetta, Peter Fazekas, Santiago Echeverri, Jason Leddington, Federico Lauria, for their great feedback and support.

I am more than grateful to the Centre for Philosophy of Memory and its members, and among them I am especially thankful to Andre Sant'Anna, Chris McCarroll, Kourken Michaelian, Denis Perrin.

I owe a lot to an informal group of philosophers in Lyon and Paris, and in particular to Quentin Bazin, Jean-Felix Gross, Vincent Beaubois, Daphné Le Roux, Camille Chamois, Mona Gérardin-Laverge. They have been here when I needed help to rehearse, for advice, for review. For their unvaluable support and friendship, I am more than grateful.

I thank the students I had both in the University of Paris Nanterre from 2014 to 2016 and in the University of Grenoble Alpes in 2019, for their motivation and interesting discussions.

I am thankful for the feedback I received on several occasions when I presented my work in Conferences these past years.

I am also grateful to the artist Armin Zoghi for a great collaboration.

I am also really grateful to Vicki Velsor, for her proofreading work on the chapters of my thesis.

In addition, I am also grateful to the people that have helped me in many ways to accomplish this work, with advice, friendship, material and psychological support these last years.

Among them, I am especially grateful to Catherine Gérardin, Thierry Laverge and Mona Gérardin-Laverge who have trusted me for many years, and for our shared history.

I am grateful to Marcelle Gérardin and her confabulations, to Denise and Robert Laverge, to Alice Caspi and her friends, to Marion Laverge, Fanny Caspi, Elaia, Ludmilla.

Among them, I am also grateful to the members of the Chateau Lazer, and especially to Coline Bassenne, Quentin Bazin Baudrand, Jean-Felix Gross, Marion Lechevallier, Robin Lhuillier, Elsa Sabater, Benjamin Suchet, Till Panfiloff.

Among them I am also grateful to the members of Proudhon: Mona Gérardin-Laverge, again, Hugo Jarry, Logan de Carvalho, David Amalric, Margaux Vermerie.

Among them I am also grateful to the members of the Self-help group at Lieues and especially to Marie Papon.

Among them, I am more than grateful to Zoé Poutrel, Cécile Box, Marion Couzinié, Pauline Buttner, Lucile Froittier, Sophia Achhibat, Coline Pascal Gemen, Céline Auclair, Maroussia Kroichvili, Maxime Pambet, Abel-Antoine Vial, Thibault de Raymond, Raymond IV, and RaymonDe, Gus Sauzet, Paul Ruellan, Hugo DallaCosta, Quentin Thirionnet, Nicolas Mollard, Florian Bardet, Jessica Jargot, Thierry Jolivet, Guillaume Kervevan, Benoit Martin, Lucas Delesvaux, Julien Cavagnis, Nicolo Sbuelz, Milan Otal, Julien Derivaz, Manuel Petitdemange-Bodin, Magali Pineau, Gabriel Lechevalier, Julien Reynier, Marilou, Laurent, Paloma, Vivianne, Nours, Sabine, Julien, Pavel et Marin, Laetitia de Villepin and Eric Vernier; Paulette Kayser and Guy Hild, Sylviane Piselli, François Lebrun, Elisabeth Gemen, François-Xavier Pascal, Agnan Kroichvili, Martine Melin, la rue Pêcherie, Laurence, Anne Allombert, Claire Thouvenot, Perrine Caline, Tristan Josset, Vincent Weber, Viviana Lipuma, Claire Chassot, Marion Geoffrey, Lia, Ugo Vitu,

Sandro, Quentin Harada, Romain, Quentin, Valentin, William, Michel Bazin, Mireille Baudrand, the extended Plat du Loup and Eric Soudan, Léo Collier and Léo, Chloé Giraud, Caroline Pénicaud, Flo Morin, Ana Rita, FX, Jean Baptiste Cognet, Joseph Paris, Michel Lepesant, Djénane Boyer-Chatenet, and all the others who have contributed in one way or another.

Table of Contents

RESUME.....	4
ABSTRACT.....	5
AFFILIATIONS.....	7
RESUME SUBSTANTIEL.....	9
PREMIERE PARTIE.....	10
PREMIER CHAPITRE.....	10
SECOND CHAPITRE.....	14
SECONDE PARTIE.....	18
TROISIEME CHAPITRE.....	19
QUATRIEME CHAPITRE.....	23
TROISIEME PARTIE.....	27
CINQUIEME CHAPITRE.....	28
CONCLUSION.....	33
ACKNOWLEDGMENTS.....	36
TABLE OF CONTENTS.....	43
INTRODUCTION.....	47
SITUATION.....	47
PROBLEMS AND DIRECTIONS.....	49

METHOD	50
OUTLINE.....	50

FIRST PART – MEMORY AND PERSONAL IDENTITY 52

CHAPTER 1 – AGAINST CIRCULARITY	53
INTRODUCTION.....	53
SECTION 1 – THE THREE CLAIMS	55
SECTION 2 – THE CIRCULARITY OBJECTION	67
SECTION 3 – AGAINST CIRCULARITY: SELF-IDENTIFICATION WITHOUT PRIOR PERSONAL IDENTITY	70
SECTION 4 – CONCLUDING REMARKS.	78
CHAPTER 2 – AGAINST THE PRESERVATIVE VIEW OF MEMORY	81
INTRODUCTION.....	81
SECTION 1 – LOCKE AND MEMORY AS A STOREHOUSE	84
SECTION 2 – BERKELEY’S AND REID’S OBJECTIONS	93
SECTION 3 – LOCKEAN RESPONSES.....	97
SECTION 4 – A HYBRID VIEW OF EPISODIC MEMORY	107
SECTION 5 – BACK TO THE OBJECTION	119
SECTION 6 – CONCLUDING REMARKS	120

SECOND PART – THE CONSTRUCTIVE DIMENSION OF EPISODIC MEMORY 124

CHAPTER 3 – ARE FALSE MEMORIES MEMORIES? THE CONDITIONS FOR REMEMBERING.....	125
INTRODUCTION.....	125
SECTION 1 – A CASE.	127
SECTION 2 – THE FACTIVITY CONDITION AS A NECESSARY CONDITION FOR REMEMBERING	132
SECTION 3 – THE CAUSAL CONDITION AS A NECESSARY CONDITION FOR REMEMBERING	152
SECTION 4 – CONCLUDING REMARKS	165
CHAPTER 4 – NON-EPISTEMIC MEMORY EVALUATIONS AND NON-EPISTEMIC MEMORY FUNCTIONS.167	
INTRODUCTION.....	167
SECTION 1 – EPISTEMIC MEMORY AND EMPIRICAL MEMORY AS DISTINCT	168
SECTION 2 – EPISTEMIC EVALUATIONS OF EPISODIC MEMORY DEPEND ON CONTEXTS	172
SECTION 3 – MEMORY SCIENCE LEADS TO REVISING EPISTEMIC MEMORY EVALUATIONS.	175
SECTION 4 – NON-EPISTEMIC FUNCTIONS OF EPISODIC MEMORY	181
SECTION 5 – CONCLUDING REMARKS	185

THIRD PART – RETHINKING MEMORY TO REDEFINE PERSONAL IDENTITY. 188

CHAPTER 5 – PERSONAL IDENTITY, CONSTRUCTIVE MEMORY AND NARRATIVITY.....	189
INTRODUCTION.....	189
SECTION 1 – THE MISSING PIECES OF MEMORY THEORIES	191
SECTION 2 – MACINTYRE’S NARRATIVE VIEW.....	195
SECTION 3 – INCOMPLETENESS OF MEMORY AND NARRATIVE THEORIES – THEIR COMMON PRESUPPOSITIONS.	199
SECTION 4 – AGAINST NARRATIVITY	200
SECTION 5 – AN ATTEMPT TO REVISE THE MEMORY THEORY OF PERSONAL IDENTITY: A CONSTRUCTIVE MEMORY THEORY.....	206
SECTION 6 – CONCLUDING REMARKS.	214

CONCLUSION	217
MEMORY AND PERSONAL IDENTITY IN LOCKE	218
PERSONAL IDENTITY AND CIRCULARITY	218
MEMORY AND THE PRESERVATIVE VIEW	221
THE CONSTRUCTIVE DIMENSION OF EPISODIC MEMORY	226
THE CONDITIONS FOR REMEMBERING.....	226
NON-EPISTEMIC FUNCTIONS OF EPISODIC MEMORY	230
CONSTRUCTIVE MEMORY AND PERSONAL IDENTITY	234
<u>BIBLIOGRAPHY</u>	<u>242</u>
<u>FULL TABLE OF CONTENTS</u>	<u>262</u>

Introduction

Situation

There is something about myself that seems obvious: I am not reborn at each new moment. I am born once, somewhere and sometime. It is not supposed to happen again any time or anywhere. The day I die, it will be over. My life as a living human being will end, as it started: dramatically. I thus have one special beginning and one particular end and belong to a particular time and place.

Another weird obvious thing is that I know this. I can recognize myself; most of the time, I do not doubt who I am. Even though I have changed a lot since I was a child, I know I still am myself. I was myself then, yesterday and I am myself now, I am the same person.

There are two temporalities of the question of personal identity: what grounds the self synchronically, at t_1 ? And what grounds diachronic personal identity? i.e. what are the conditions for two episodes to be co-personal? The notion of self seems to

be concerned with the synchronic question and the notion of personal identity with the diachronic question of self-extension through time. This latter question is the one I will focus on.

Furthermore, there are also two dimensions of the question about myself: a metaphysical one: I am a someone; an epistemic one: I know who I am. And, at first glance, the answer to the epistemic one seems to depend on the answer to the metaphysical one. To recognize myself, I need to be a self. Self-knowledge seems to be merely the acknowledgement of something that already is objectively.

Those two dimensions of the question have their source in the philosophy of John Locke, 1694. In the *Essay Concerning Human Understanding*, II, xxvii, he proposes a theory of personal identity. What his theory exactly is has been a matter of debate up to the present day. According to a widespread reading of Locke, diachronic personal identity depends on memory. Being a person is to recognize oneself thanks to the recognition of the same consciousness between the rememberer and the remembered subject through a memory. Diachronic personal identity depends metaphysically and epistemically on memory, because memory allows one to extend through time, and to have a temporal depth.

Memory is a mental capacity that has long been thought of as a preservative capacity. The metaphors of memory in the history of philosophy are almost always varieties of storehouse metaphors (Draaisma, 2000). Memory is supposed to store and record information in order to retrieve it accurately.

As such, memory would play an important epistemic role for personal identity. It allows individuals to access the truth of their past.

Problems and directions

However, it seems to me that the question of personal identity is not yet solved. It seems to me that the concept of memory is not clear and that the role that memory can play for personal identity still awaits clarification.

First, Locke's work on personal identity, as I understand it, seems to go beyond the dichotomy between the metaphysical question and the epistemic question of personal identity. Persons are not substances that could be objects of knowledge, but their relation to themselves produces something about themselves. The answer to the metaphysical question depends in fact on the answer to the epistemic question in Locke: how is this even possible?

Second, although memory has been long considered as a storehouse, interestingly, since the late 19th, psychologists have explored its constructive dimension. The picture of memory and in particular of episodic memory (i.e. memory for episodes, Tulving, 1985) that emerges from empirical research is the picture of a constructive capacity both at encoding and at retrieval.

Third, if memory has a constructive dimension, how does that change the question of personal identity?

Fourth, the field of philosophy of memory has emerged recently (Bernecker & Michaelian, 2017). And within this field an important question is to determine what the functions of episodic memory are. Is it only to know the past? If memory has a constructive dimension, how can it be a source of knowledge of the past? How can we distinguish between genuine remembering, misremembering and confabulating? Should we maintain a distinction between them?

Those four lines of questions are at the basis of this research. At stake is the attempt to better understand personal identity with an empirically informed concept of episodic memory.

Method

This thesis is grounded in history of philosophy, and in particular in Lockean studies. I rely on textual evidence to discuss interpretations of Locke's theory. It also takes a stance in a philosophical tradition that explores the question of personal identity, and I favor a discussion on psychological continuity and on the epistemic dimension of the diachronic question of personal identity.

It also is concerned with the history of psychology, and philosophy of psychology. In particular, I retrace the emergence of the research on the constructive dimension of episodic memory. This research is important for an empirically informed philosophy of mind, since I am interested in mental capacities, and I discuss and rely on empirical arguments coming from memory science. Finally, there are strong connections between empirical research on memory and the new field of philosophy of memory, and I take a stand in several discussions that are happening in this field, about the definition, the functions and the underlying mechanisms of episodic memory.

Outline

In the first part of the thesis, I start with an exploration of John Locke's memory theory of personal identity. In the first chapter of the thesis, I attempt to overcome the dichotomy between epistemic theory and metaphysical theory of the self. I argue that to be a person, in Locke's view, is to recognize oneself as the same at different moments of time and thus, in this act of self-recognition, to constitute oneself as a person with a temporal extension. The answers to the epistemic and to the metaphysical question are not distinct. I discuss the famous objection of circularity, and I propose an empirical argument against it. I compare the Lockean case of fatal error and the case of false memories studied in the psychological literature, when someone remembers an episode she didn't live. I argue that recognizing oneself does not presuppose personal identity but can constitute it. In the second chapter,

following Reid, and grounding on empirical research about episodic memory, I give arguments against Locke's preservative view of memory. My point is that the concept of memory has to be revised to play a grounding role in personal identity and to be consistent with what is known today about this capacity. I show that episodic memory has a constructive dimension and sketch a hybrid view of memory.

In the second part of the thesis, I deepen the reflection on the concept of memory. This reflection takes part in a contemporary exploration in memory science and in philosophy of memory about the conditions of remembering and about the functions of episodic memory. In chapter 3, I argue that neither the factivity condition nor the causal condition should be considered as necessary conditions for remembering. In chapter 4, I discuss the functions of episodic memory and contend that episodic memory has non epistemic functions, even though they are rarely considered in the contemporary literature. Episodic memory has functions related to psychological continuity and potentiality for action.

In the last part of the thesis, I propose a constructive memory theory of personal identity. I argue that a constructive memory can play a double role: epistemic and metaphysical. It is at the same time a capacity which allows me to recognize myself and, because this recognition is not a simple recognition of an object which precedes it, it can constitute me, as a person with a history, capacities and intentions. I contend that the question of the truth of memory as well as the truth of the self needs to be transcended. They can in fact be secondary in a reflection on personal identity. In this regard, I argue that that self-recognition, regardless of accuracy, is crucial for psychological continuity. Therefore, I claim that it might be useful to change our practices towards people who confabulate. Rather than focusing on fidelity to the past and rejecting confabulations for what they cannot do, it might be useful to focus on what confabulations allow people to do, i.e. to build a sense of psychological continuity. In this regard, the memory theory that I propose allows us to think that people who remember inaccurately, or even confabulate, can recognize their personal identity and thus constitute themselves.

First Part
– Memory and personal identity

Chapter 1

– Against Circularity

Introduction

In philosophy, there are two parallel traditions of reflection on selves and persons (Schechtman 2016). One of them is the epistemological tradition and focuses on the possibility of self-knowledge, on the transparency of self-access, and on the different levels of self-knowledge (Neisser, 1988; Fernández, 2013; Perry, 2017). The other one deals with the metaphysics of personal identity, that is, the nature of persons¹. It is a reflection on the necessary and sufficient conditions for personhood, and for the identity of persons, namely on the conditions under which two episodes are episodes of the same person's life (Perry 2008; Williams 1973; Olson 2017). These two traditions neither ask the same kind of questions nor have the same object: the self seems to be the privileged object of epistemology and the person the object of metaphysics². However, they both have their source in the philosophy of John

¹ On the range of different questions about personal identity, see (Olson 2017).

² I thus agree with Shelley Weinberg's interpretation of those terms in Locke. 'I interpret the difference to be that 'self' denotes that which we perceive from the subjective (first-personal) point of view, where 'person' denotes the same thing as referred to from the objective (third-personal) point of view.' (Weinberg 2012, p. 3). For a discussion on the different interpretations

Locke (Locke 1694). In *The Essay*, II, xxviii³, Locke asks both kinds of questions. According to him though, they are inseparable, and nonetheless prioritized. The answer about personal identity (what are the conditions of personal identity?) depends on an answer about the means of self-access (how can a person identify herself as the same person in different times and places and be capable of self-ascription of past and future episodes of her own life?): in Locke's account, "the definition of the self is correlated to the mode of its recognition" (Forest 2018).

Beyond academic philosophy, there is a widespread idea that memory is the ground of our personal identity. In several ways: (1) memory is a capacity that allows one to acknowledge her personal identity (one example of this belief is summed up in Linda Grant's book title: *Remind me who I am, again* (Grant 2011)); (2) this capacity is a necessary condition of personhood; being the same person is to recognize myself as the same person (as noted by Craver 2012, who gives objections to this view); and (3) my identity is my memory (versions of this claim can be found in lots of introductions to psychological books on human memory such as Loftus 1980; Schacter 2008; Eustache and Desgranges 2015).

These three claims also have their source in a popular reading of John Locke's theory of personal identity as a memory theory. In this paper, I examine both the epistemic and the metaphysical questions about personal identity and the three related claims.

I argue (1) that Locke, as an answer to those two questions, has defended three claims about personal identity:

Epistemic claim: Cognitive capacities that allow self-recognition epistemically ground diachronic personal identity.

Reduction claim: Being a person is to have and exercise psychological capacities which allow one to recognize oneself.

of the difference, see [Weinberg 2012](#), footnote 5. See Thiel 2011 and Yolton 2004 for alternative views.

³ I will refer to the *Essay* in this manner: E, BOOK, chapter, section.

Metaphysical claim: Personal identity is metaphysically constituted by the epistemic relation one has with oneself.

The reduction claim has been regarded as a weak claim. I show that one of the most famous objections his theory has received is directed towards the reduction claim: the circularity objection – self-recognition (through memory or consciousness or both) cannot define personal identity since it presupposes it (Butler 1736) (2).

I argue that the reduction claim states that personal identity depends on cognitive capacities and on an epistemic and constitutive relation one has with oneself thanks to these capacities. I contend that this is the strongest Lockean claim, and that it is consistent with Locke's ontological views regarding the ideas of relation. Eventually, I give an empirical argument against the circularity objection. I envision cases where same-consciousness or auto-noetic consciousness (Tulving 1985) happens without prior personal identity. In particular, I give an analysis of false memories where self-identification happens without prior personal identity between the remembered and the rememberer (3).

Section 1 – The three claims

In this section I rely on textual evidence and propose a reconstruction of Locke's argument to show that Locke's theory of personal identity aims at answering both the epistemic and the metaphysical questions of personal identity. I contend that Locke has explained personal identity by the means of its apprehension. His theory is composed of three related claims. Here are the three claims in slightly more detail:

Epistemic claim:

a. Synchronic aspect – I can recognize myself as an individuated person thanks to self-consciousness, present in any of my thoughts.

b. Diachronic aspect – I can recognize myself as the same person extended through time thanks to the sameness of consciousness present both in my present thought and in my memory. When I remember an action or thought I access two facts: *I* am remembering, and I am remembering something *I* was conscious of.

Reduction claim:

Being a person is to have and exercise psychological capacities which allow one to recognize oneself both synchronically and diachronically. Locke reduces the metaphysical self to the means of its recognition by a subject. Thus, he makes a leap from the epistemic question of personal identity to the metaphysical question and reduces the metaphysical definition of personal identity to its epistemic apprehension.

Metaphysical claim: Personal identity is metaphysically constituted by the epistemic relation one has with herself. Synchronically, being a self is related to self-consciousness as a reflexive capacity. Diachronically, being the same person is to recognize myself as extended through time thanks to the recognition of the same consciousness.

1.1. Epistemic claim

In the *Essay*, Locke gives an answer to the epistemic question of how one can recognize oneself as a self (synchronically) and as the same person (diachronically): self-consciousness allows self-recognition and thus epistemically grounds personal identity.

First, in *The Essay*, II, xxvii, he defines identity and diversity as ideas of relation⁴:

“That therefore that had one beginning is the same thing, and that which had a different beginning in time and place from that

⁴ The relational account of personal identity has been criticized, in particular by Reid (1785, in Perry, 2008) and Butler (1736). While it may be so according to Locke, the fact that identity could be considered as a relation is subject to controversies (see for instance Geach 1967).

is not the same but divers. That which has made the difficulty about this relation, has been the little care and attention used in having precise notions of the things to which it is attributed” E, II, xxvii, 1.

He argues that ideas of relation come by comparison of two or more ideas:

“The last sort of complex Ideas, is that we call Relation, which consists in the consideration and comparing one Idea with another.”

E, II, xii, 7.

As a consequence, the idea of identity comes by comparison of two or more ideas:

“Another occasion, the mind often takes of comparing, is the very beginning of things, when considering any things as existing at any determined time and place, we compare it with it self existing at another time, and thereon form the ideas of identity and diversity”

E, II, xxvii, 1.

Second, Locke differentiates between the identity of a human and the identity of a person⁵: according to him, a ‘man’ is an individual whose identity consists in his life, attached to a specific body. But a ‘person’ is a self-conscious entity. This distinction is crucial, and I will come back to it.

“It being one thing to be the same substance, another the same man, and a third a same person, if person, man, and substance, are three names standing for different ideas; for such as is the idea belonging to that name, such must be the identity.” E, II, xxvii, 7.

If Heliogabalus is changed into a pig, he is no longer a man, but he can still be a self, and if he can identify with Heliogabalus, he can be the same person

⁵ There is a debate about the equivalence between ideas of things and things in Locke. And some scholars have recently defended that they were not equivalent (see for instance Lennon 2001). Here, though, I use ‘idea of identity’ and ‘identity’ interchangeably.

as Heliogabalus (II, xxvii, 6). If a parrot is self-conscious it cannot be a man but can be a person (II, xxvii, 8).

The idea of a human stands for an individual with a specific body:

“For I presume it is not the idea of a thinking or rational being alone, that makes the idea of a man in most people’s sense; but of a body so and so shaped joined to it”. E, II, xxvii, 8.

The idea of a person is attached to mental capacities. The properties of the person are self-consciousness, reflexivity, and rationality. Person stands for a thinking, intelligent being capable of reason, reflection and self-recognition:

“we must consider what person stands for; which, I think, is a thinking intelligent being, that has reason and reflection, and can consider it self as it self, the same thinking thing in different times and places...” E, II, xxvii, 9.

Consciousness is the mean of synchronic self-apprehension:

“... which it does only by that consciousness...” E, II, xxvii, 9.

And thinking entails self-consciousness:

“...which is inseparable from thinking, and as it seems to me, essential to it: it being impossible for anyone to perceive, without perceiving, that he does perceive” E, II, xxvii, 9.

As a consequence, the idea of myself comes synchronically by consciousness and is present in any of my thoughts. Thus, consciousness gives me an epistemic access to myself in the sense that it allows me to identify as the owner of my thoughts⁶, actions and perceptions.

The self is thus epistemically and synchronically accessed by self-consciousness:

⁶ This aspect of Locke’s theory has been emphasized in particular by Winkler 1991 who claims that Locke’s theory of personal identity is a theory of action and thought appropriation.

“When we see, hear, smell, taste, feel, meditate, or will anything, we know that we do so. Thus it is always as to our present sensations and perceptions: and by this every one is to himself that which he calls self” E, II, xxvii, 3.

When one perceives the world, one is conscious of one’s own perceptions, and conscious that one is the subject of one’s own consciousness. Thus, perceiving entails an epistemic access to oneself in the sense that it allows one to apprehend oneself as a subject of sensation and perception (Flew 1951).

Third, in Locke’s account, I can have access to a past action or experience by remembering⁷ (E, II, x). A memory is a past-oriented thought characterized by a reference (to a past event), a sense of ownership (I am remembering), and a specific phenomenology: the past-oriented thought appears to me as a recollection of an experience I have already had – remembering is thus, according to Locke, accompanied by what recently has been called *autonoetic consciousness* (Tulving 1985). Namely consciousness “that mediates an individual’s awareness of his or her existence and identity in subjective time extending from the personal past though present to the personal future. It provides the characteristic phenomenal flavor of the experience of remembering” (Tulving, 1985, p. 1), and is the type of consciousness attached to episodic memory, namely, the memory for personal past episodes. Locke’s definition of remembering fits very well with this frame since he writes:

“so that this ‘storing of ideas in the repository of the memory’ really means only that the mind has a power in many cases to revive perceptions that it has once had, with attached to them the additional perception that it has had them before” E, II, x, 2.

The idea of the sameness of myself, extended in time, comes through the recognition of the same consciousness. As sameness is a relation, this idea comes by a comparison of a past idea of my conscious self, and my present self-consciousness:

⁷ I use remembering and memory interchangeably, because if Locke’s theory is a memory theory, it is about the retention of ideas I had, about memories and about the capacity of remembering.

“For as far as any intelligent being can repeat the idea of any past action with the same consciousness it has of its present thoughts and actions, that it is self to it self now, and so will be the same self as far as the same consciousness can extend to actions past or to come.” E, II, xxvii, 10.

Consequently, the sameness of consciousness allows me to have a diachronic access to myself and is sufficient for personal identity, as it appears in the Noah thought experiment:

“Had I the same consciousness, that I saw the Ark and Noah’s flood, as I saw an overflowing of the Thames last winter, or as that I write now, I could no more doubt that I, that write this now, that saw the Thames overflowed last winter, and that viewed the flood at the general Deluge, was the same self, place that self in what Substance you please, than that I that write this am the same my self now whilst I write (...) that I was yesterday.” E, II, xxvii, 16.

If I have a memory of seeing Noah’s Ark, i.e. a representation of an event accompanied by auto-noetic consciousness, namely with the consciousness that I in the past lived this experience I am now remembering, then, I have an epistemic access to myself. I extend to the past. Thus, memory is one of the means of self-access, and along with consciousness understood as same-consciousness, it allows one to recognize oneself and to extend to the past. It thus grounds personal identity epistemically in Locke. But not only that. In Locke, the relation between consciousness as same-consciousness and personal identity is more a relation of equivalence than a relation of foundation. In short, what has been called the "reductionism" of Locke consists in the fact that he seems to reduce personal identity to the epistemic access one has to oneself⁸.

⁸ Focusing on the epistemic access one has to oneself in Locke, I leave aside the interesting reflection on the sensitive aspect of Locke’s theory. Though for a reflection on the notion of concern in Locke, see Hamou 2004; for a reflection on the notions of sensitivity and care in Locke see Schechtman 1996, p.109 sq. See also Thiel 2015 on sensitivity.

1.2. Reduction claim

Locke makes it clear that the idea of personal identity could be sustained by an immaterial substance, but since I cannot form a clear idea of an immaterial substance – I cannot experience it – it would be a mistake (neither careful nor useful) to believe that the idea of personal identity is derived from it (E, II, XXVII, 17, 19, 23). It seems that in Locke, the mental act of self-recognition enables me to form an idea (of relation) of the sameness of consciousness extended in time, and this mental act makes the sameness of the self.

Consequently, the epistemic relation one has to him or herself is what constitutes his or her personal identity both synchronically and diachronically.

The reduction claim first appears in II, xxvii, 9:

“For since consciousness always accompanies thinking, and it is that, that makes every one to be, what he calls self; and thereby distinguishes himself from all other thinking things, in this alone consists personal identity” E, II, xxvii, 9.

What makes personal identity depends on self-access and on consciousness. So, the reduction claim seems to be first a claim about synchronic personal identity. But the reduction claim also concerns the diachronic self:

“And thus, by this consciousness, he finds himself to be the same self which did such or such an action some years since, by which he comes to be happy or miserable now. In all which account of self, the same numerical substance is not considered, as making the same self: but the same continued consciousness, in which several substances may have been united, and again separated from it, which, whilst they continued in a vital union with that, wherein this consciousness then resided, made a part of that same self.” E, II, xxvii, 25.

Finding myself as being the same self, also makes myself as the same self, and makes my personal identity. As Forest, 2018 puts it, Locke reduces the self to the

means of its own recognition. Consciousness allows diachronic self-recognition and self-constitution. Personal diachronic identity is made by the same consciousness.

1.3. Metaphysical claim⁹

Consequently, the metaphysical claim which follows the reduction claim holds that personal identity is grounded on capacities of self-recognition:

“Consciousness makes personal identity”, E, II, xxvii, 10.

1.3.1. Consciousness and/or memory?

If scholars agree on the importance of memory for the epistemic dimension of personal identity, there is a disagreement on the role of memory for the metaphysics of persons. Did Locke defend a memory theory of personal identity from both an epistemic and a metaphysical point of view? The idea that he did is the canonical interpretation and has received defenses in recent literature (Stuart 2015; Garrett 2003), but it is the matter of a lively debate. Atherton, 1983 has famously argued against this standard view. According to her, the standard view (1) interprets Locke as having defended a memory theory of personal identity, namely, that what makes a person a same person despite changes, is the phenomenal feature of remembering. (2) The standard view assumes that Locke wanted to build a notion of personal identity to ‘serve essentially practical ends’ (Atherton 1983, p. 273), namely, to decide who is responsible for what. But, according to Atherton, the standard view is misleading: the essential ingredient of Locke’s theory of personal identity is consciousness, and not only is self-consciousness obviously distinct from memory, but same consciousness also is.

⁹ If several passages of the *Essay* (such as *Essay*, II, xiii, 16, for instance), can be interpreted as showing Locke’s indifference towards ontological questions (Yolton, 2000, p. 50), Hamou and Pécharman have recently shown that natural philosophy and metaphysics were crucial in Locke’s text (Hamou & Pécharman, 2018, p. 11).

Against the memory interpretations there are two main lines of argument. One focuses on the distinction between same consciousness and memory; the second focuses on the concept of memory in Locke.

Some of Locke's commentators take the memory interpretations of Locke's theory as being guilty of a confusion between same-consciousness and memory. Strawson, 2011, for instance, in chapter 9 of his book offers a reflection on how consciousness and memory are distinct in Locke; he relies in particular on E, II xxvii, 9, 16, 26 to claim:

"In each case the basic reference point for attribution of Consciousness is the subject of experience considered in the present moment. Consciousness cant be identified with memory."
(Strawson, 2011, p. 74).

Strawson rejects the idea that consciousness of the past is the same as memory. Atherton, 1983, Weinberg, 2012, Hamou, 2014 also reject the idea that Locke equates same consciousness and memory. I will come back to the distinction between memory and same-consciousness in Chapter 2, and show that even though Locke distinguishes between them in some occasions, in some other ones, he doesn't. Though, I believe that the reading I propose (which gives an important place to memory in Locke's theory) holds regardless of whether there is a strong distinction between memory and same consciousness or whether they are equivalent. Here, I give an answer to the line of arguments against memory interpretations on the basis of the distinction between memory and same-consciousness, following (Garrett 2003). Garrett, for instance, holds a memory interpretation but acknowledges the distinction in Locke between memory and same consciousness. According to him memory is not equivalent to but sufficient for same consciousness. Memory gives rise to same-consciousness, and furthermore is sufficient for same consciousness. Since every thought or perception in Locke's sense involves consciousness, every memory is also a conscious memory. Since memory and consciousness, as noted by Garrett involve "a representation of the self as part of their content" (Garrett, 2003, p. 12), when remembering, the subject accesses at least two representations of the self (past and

present). Memory is thus sufficient for some consciousness between a past and a present self. And some consciousness is sufficient for personal identity extended to the past, at least from an epistemic point of view. Locke's account thus crucially involves memory.

But what kind of memory can play this role? According to Atherton, the standard view is misleading in how it characterizes Locke's conception of memory:

"he never did, in fact, hold a memory theory in this phenomenalist sense" (Atherton, 1983, p. 274).

The idea that it is a mistake to believe that Locke's notion of memory is merely phenomenal is also endorsed by Garrett. He argues that memory is not only phenomenal memory or seeming memory in Locke. According to him, in Locke, memory involves the revival of an earlier perception, made possible by "the activation of traces causally derived from this earlier perception" (Garrett, 2003, p. 18).

I think that this view is related to a widespread intuition according to which seeming memory is not memory, and genuine memory should be distinguished from mere phenomenal memory. I see three problems with this view. First, Locke cares about phenomenal memory, as made evident in most of the thought-experiments he proposes, when one remembers being Noah, or Socrates, or the cobbler. If I remember seeing Noah's Ark, as a subject of experience, and recognize the same consciousness in this past perception and in my present thought, I extend myself to this past event. Second, it is generally assumed that Locke endorsed a causal theory of memory, according to which remembering depends on causal traces originated in the past experience, but he does not make it clear. Third, as I shall argue in more detail in section 3, Locke's crucial passage on fatal error entails that if a subject wrongly remembers a past experience, she nonetheless will identify with it, and thus will be the same person as the remembered person, at least to herself.

1.3.2. An objective fact?

The claim according to which, for Locke, personal identity depends on a form of self-knowledge and on a first-person access, has received critiques among Lockeans in recent literature. Hamou, 2014 and Weinberg, 2012 have given two detailed critiques of such a view. Both defend an objectivist thesis about Locke's theory of personal identity. Their views are highly interesting because they have the advantage of overcoming traditional objections against Locke, and second, they specifically solve the problem of error or self-deception. Subjects are sometimes wrong about themselves, so how could their personal identity depend on their fallible self-access? How could Locke have defended such a view? According to Weinberg 2012¹⁰, Lockean consciousness is not only a means of self-recognition but also is an objective fact. To her, Locke's view implies consciousness as a metaphysical fact, inaccessible to the subject. According to Hamou, 2014, personal identity in Locke depends on something objective which is 'the same train of conscious thoughts'. They believe that there is in Locke's theory something like an objective fact, and that the "cognitionist" view of person (Hamou 2014) relies on a misunderstanding. Persons can be wrong about themselves; their personal identity cannot depend on knowledge or epistemic capacities. If Hamou and Weinberg are right, that would save Locke from the problem of error. Though, I am not sure that this is totally consistent with some passages of Locke's chapter on identity and difference. In several places, Locke insists on the necessity of a first-person access

¹⁰ According to Weinberg, we can find in the Essay, II, XXVII, evidence that "consciousness (the I, self, or person) is something that persists through our momentary conscious states of ourselves". She writes: "He talks about there being "two distinct incommunicable consciousnesses" (E, II, XXVII, 23) alternating by day and night, (...) "the same consciousness being transferr'd from one thinking substance to another" (E, II, XXVII, 13). "same consciousness extended" (E II XXVII, 9, 10, 16), "reaching" (E II XXVII, 9, 17), and "continued" (E II XXVII, 25) into the past or the future" (Weinberg, 2012). In her vocabulary, there is thus a metaphysical fact of consciousness in Locke's account: "Locke seems to be saying that despite our inability to penetrate to the real metaphysics of the external world or even our own minds, we have experience of a continually existing consciousness. Any being capable of happiness and misery "must grant," he says, that there is something enduring that he is aware of as himself and there is something for which he is concerned and wants to see happy rather than miserable." (Weinberg, 2012, p. 8) and, Weinberg 2012, p. 13: "With respect to the continuing consciousness we can say that just as an objectively existing world is present and certain to me in sensitive knowledge, an objectively existing ongoing self is present and certain to me in virtue of being conscious of myself as perceiving ideas."

for the existence of the self or the person. For instance, in E, II, xxvii, 24, Locke writes:

“if there be any part of its existence, which I cannot upon recollection join with that present consciousness, whereby I am now my self, it is in that part of its existence no more my self, than any other immaterial being.” E, II, xxvii, 24.

In this passage Locke is clear: being myself (and not only recognizing myself) is impossible without a capacity to recognize myself. More precisely, and interestingly, in the quote above, Locke argues that in the detail my personal identity depends on my access to it in a first-person perspective. Namely, for an episode to be an episode of my personal life (and not of my human life), I have to be capable of joining it with my present consciousness.

Not only does self-access seem essential to Locke’s theory of personal identity, but epistemic self-access is essential. In his notes on his copy of Sergeant’s book, Locke writes:

“A man has the individuality of a man before he has Knowledg, but he is not a person before he has Knowledg” (Locke & Sergeant, 1697, p. 265), also cited in (Thiel, 2011, p. 202).

As it appears here, and as noted by (Thiel 2011), there is something crucial in Locke’s distinction between the identity of a man and the identity of a person. According to Locke, a person is a person when she has knowledge of herself. It means that personal identity depends primarily on one’s epistemic access to oneself. It also means that personal identity is not like the identity of human individuals. The identity of human individuals can be objectified from a third person perspective. The identity of persons cannot, but always depends on a first person-perspective¹¹.

¹¹ I here use the vocabulary of first-person perspective and third-person perspective as it used in Lockean literature to distinguish between objective and subjective facts. However, this vocabulary might be misleading since it might seem to presuppose that I can only have first-person perspectives on myself. Though, against this view, (McCarroll 2018) has shown that one can remember and experience events from an unoccupied point of view.

This thus leads to two major issues. One is the circularity problem: how can self-access ground personal identity since it presupposes a self or a person? A second related issue concerns the problem of error of identification. I discuss them in the following sections.

Section 2 – The circularity objection

The circularity objection was in the air in Locke's time. It had been raised against Sherlock's notion of personality by South in the early 1690's. Then, from the late 1690's it was raised against Locke, by (among others--for a review, see Thiel, 2011, p. 196 sq. ; Sergeant 1697; Lee 1702) , and finally raised more famously by Butler 1736. The core idea is thus: consciousness presupposes a self capable of it, and thus Locke's definition of the self by self-consciousness is circular. Sergeant writes:

“Tis evident that the Individual Thing must, (in priority of Nature or Reason) be first constituted such, ere it can be capable of Existence. Wherefore 'tis impossible that Existence, consider it how we will, can be in any manner the Principle of Individuation, the constitution of the Individuum being presupposed to it.”
(Sergeant, 1697, p. 260).

To Sergeant, as explained by (Thiel, 2011, p. 192 sq.), Locke's view is circular in two regards. First, it presupposes an individual capable of existence, thus its existence cannot constitute it. Second, consciousness cannot play a role in the individuation of selves, rather it is only the means by which we know ourselves. But Individuality is necessarily antecedent to consciousness. He writes:

“Our Person, or Individual Self ... is the Object that Consciousness; and Objects must be antecedent and presupposed to the Acts which are employ'd about them, because the Objects are the Cause of those Acts.” (Sergeant, 1697, p. 267).

Persons are the cause of both their mental and physical acts, and thus cannot be constituted by those. Consciousness, as a mental act, presupposes personal identity. Sergeant thus rejects the diachronic reduction claim, and therefore the metaphysical claim.

In a similar trend, Butler's (1736) version of the circularity objection is the most famous and is thus:

“[3] But though consciousness of what is past does thus ascertain our personal identity to ourselves, yet, to say that it makes personal identity, or is necessary to our being the same persons, is to say, that a person has not existed a single moment, nor done one action, but what she can remember; indeed none but what she reflects upon. And 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.

[4] This wonderful mistake may possibly have arisen from hence, that to be endued with consciousness, is inseparable from the idea of a person, or intelligent being. For this might be expressed inaccurately thus, that consciousness makes personality; and from hence it might be concluded to make personal identity. But though present consciousness of what we at present do and feel, is necessary to our being the persons we now are; yet present consciousness of past actions, or feelings, is not necessary to our being the same persons who performed those actions, or had those feelings¹².” (Butler 1736)

Butler's argument is focused on consciousness. He seems to accept the synchronic part of Locke's account since he states: “present consciousness of what we at present do and feel, is necessary to our being the persons we now are”. Hence, he

¹² I will focus on the well-known circularity objection and not on the last claim of [4]. It is similar to Berkeley's and Reid's objection of failing transitivity (Berkeley 1732) (Reid 1785) and has received answers elsewhere, for instance see Bolton, 1994, (in Barber & Gracia, 1994) (Quinton 1962), in (Perry 2008). But I shall discuss the issue of transitivity elsewhere.

accepts the leap from an epistemic definition to a metaphysical one, concerning the synchronicity of consciousness. As he says that “consciousness of what is past does thus ascertain our personal identity to ourselves”, he also seems to accept the epistemic part of the diachronic answer to the question of personal identity.

Butler’s circularity objection (“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”) is directed towards the claim that same consciousness could define personal identity. Like Sergeant’s, his criticism aims at the diachronic reduction claim: he refuses to accept the leap from the epistemic definition to the metaphysical one concerning diachronic identity of persons. Perrin 2017 has shown that there are two main escape routes from the circularity objection in the neo-Lockean literature. One is to deny that Locke has grounded personal identity on phenomenal memory (Atherton 1983; Garrett 2003; Gordon-Roth 2019; Gustafsson 2010; Hamou 2014; Weinberg 2012b, 2016; Yaffe 2007). I have already raised concerns against these views. The other one is an attempt to save the importance of phenomenal memory in the constitution of personal identity and proposes that memories do not have to be personal to ensure psychological continuity; in this view quasi-memory could suffice (Parfit 1984; Shoemaker 1970). Parfit’s and Shoemaker’s quasi-memory views rely on a causal theory of memory. For a subject to q-remember an experience, the experience must have happened and be the cause of the subject’s q-memory. In a sense quasi-memory still presupposes that memories are faithful pictures of past events. My proposal is compatible with Parfit’s and Shoemaker’s views according to which what matters for personal identity is psychological connectedness but is incompatible with the preservative view that the quasi-memory account presupposes. As shown by recent literature in psychology and neuropsychology (see for example Addis et al. 2012), memory is not only a preservative capacity but also has a constructive dimension. I will come back to this issue in more detail in Chapter 2.

Another approach consists in insisting on the distinction by Locke between Individual, Man and Person. It is the path followed by Thiel, 2011. Thiel offered a fruitful discussion of the circularity objection. According to him, it relies on a

misunderstanding of Locke's view. Consciousness does not presuppose a person but presupposes an individual capable of consciousness. If one makes sense of Locke's distinction between Individual, Man and Person, the problem of circularity is weak¹³. As said earlier, I also believe that the distinction between the notions of human individual and person is crucial in Locke. As I understand Locke's text, one can be wrong about one's individual life but cannot be wrong about one's personal identity. Personal identity depends on self-apprehension.

In what follows, I propose an empirical argument against the circularity objection consistent with my understanding of Locke's text: the epistemic access that I have to myself constitutes myself as a person and makes my personal identity. As such, personal identity depends on the means of its epistemic apprehension. Both consciousness and memory play this role in Locke's theory. Both play a constitutive role in personal identity as capacities that allow people to recognize themselves. I thus don't think that one should choose between consciousness and memory. Consciousness allows me to know myself synchronically, and diachronically, by the means of memory. Memory is not the same as same-consciousness but is a capacity that produces mental contents that allow me to self-identify.

Section 3 – Against circularity: self-identification without prior personal identity

This section proposes an answer to the circularity objection: I make a case for self-identification without prior personal identity. It goes as follows: first, I explain Butler's objection. Second, given the fact that personal identity in Locke is an idea of relation, I consider two main ways of understanding Locke's ontology of relations. I argue that both ways – for different reasons – should acknowledge the fact that personal identity depends on a mental action. Following Thiel's reading of

¹³ Thiel, 2011 argues that the transitivity problem is more difficult to overcome (I will discuss this point in the next chapter).

Locke, I argue that same-consciousness might presuppose a human individual but does not presuppose personal identity. I give an empirical argument: I envision the case of fatal errors (Locke, II, xxvii, 13) or false memories. I contend that those cases are cases where personal identity is not presupposed by self-identification and self-consciousness. Consequently, I argue that same consciousness and memory do not presuppose personal identity.

3. 1. Building the question raised by the circularity objection

"Consciousness of personal identity presupposes and therefore cannot constitute personal identity" (Butler, 1736)

What this objection means is that for personal identity to be conscious, personal identity has to be, or for self-identification to happen, prior personal identity has to be. Thus, my acknowledgment of personal identity is nothing more than the recognition of a presupposed personal identity. It could also be described thus: I self-identify in a memory, for instance, because there is already a personal identity between the subject remembered and the rememberer. Butler seems to agree with what is called now the causal theory of memory (Martin and Deutscher 1966): a memory is caused in a subject by a past experience *she* had. Consequently, self-recognition is not a supplementary fact, and cannot be said to be foundational. The identity of the rememberer and the remembered is necessary for a memory. Personal identity cannot depend upon a mental act which presupposes it.

3.2 Ontology of relations in Locke

Since personal identity is a relation of identity, the question is whether relations can be independent of a mental act or are constituted by a mental action. In Lockean exegesis, there are two main schools¹⁴:

¹⁴ Locke's ontology of relations is a matter of lively debate, and such passages as the following reveal an ambiguity in Locke's account. Anti-realist and conceptualist interpretations of Locke's ontology of relation take as textual evidence the following passages: 1. from E, II, xxv, 8: "*Relation* [is] not contained in the real existence of Things, but something extraneous, and superinduced". "On its face, this says that relations

- a. One claims that Locke is a conceptual realist, and that relations in Locke are never independent from a mental act (Gibson 1917; Perry 1967).
- b. The other one claims that Locke is a non-conceptual realist, and that relations do exist in Locke independently from their apprehension. Rickless, 2018, has proposed the most detailed defense for this claim.

For the former school, personal identity is, like any other relation, dependent on a mental act, and so consciousness of personal identity does not necessarily presuppose it since it constitutes it. The latter school, though, will have to acknowledge that Locke's treatment of personal identity is different from his treatment of other relations: first of all, person is a forensic term (and thus a conceptual entity) in Locke (E, II, xxvii, 17, 19, 20 and 22); its reality comes from a specific social and judicial sphere¹⁵. In his view, being a person is being conscious of and thus responsible for her own actions. Reward and punishment are justified if and only if they apply to my past actions, and to the past actions I can acknowledge; As Boeker 2014 has shown, "questions concerning personal identity over time are to be understood as questions concerning the existence of a subject accountability over time" (Boeker 2014:231). Namely, the concept of personhood is related to the

are not real elements of the extra-mental world but instead are creatures of the mind" (Ott, 2003, p. 160). But there are also passages where Locke seems to plead for the mental-act dependence of relations: 2. "*Mixed Modes and Relations*, [have] no other *reality*, but what they have in the Minds of Men." (E, II.xxx.4: 373), and 3. "But *mixed Modes and Relations*, being Archetypes without Patterns, and so having nothing to represent but themselves, cannot but be adequate, everything being so to itself." (E, II.xxxi.3). As (Rickless 2018) notes, these have given the basis for claiming that conceptualism "on its face, is the official position of the *Essay*" (Ott 2003, p.164). Rickless, 2018 has shown that 1, 2 and 3 refer to the ideas of relation (and not to relations themselves) such that "the reality of ideas of mixed modes and relations is therefore an intra-mental matter: such ideas are real, not because they conform to something that exists outside our minds, but because they conform to something that exists in our minds (namely, themselves)" (Rickless, 2018, p. 20). Thus, it seems that the ideas of relation, if superinduced, produce an interpretation of a phenomenon.

¹⁵ On the relations between Locke's views on personal identity and morality see Tabb 2018. She argues that Locke's use of thought experiments serves to demonstrate the usefulness of his concept of person for a science of morality.

sphere of judgement¹⁶. Second, as I have shown in section 1.1., personal identity is not treated independently from thought. It is recognized by a mental act, and a person is defined as a thinking, intelligent being capable of reason and reflection, namely capable of mental actions. In these regards, even if one thinks that Locke's general views on relations are that they can or do exist independently of a mental action, personal identity is a specific kind of relation, since it intrinsically depends on a mental action of self-recognition.

In which case, the following question arises: is personal identity necessarily presupposed by the recognition of personal identity? And the subsequent question: can recognition of personal identity happen in the absence of prior personal identity? If so, recognition of personal identity cannot presuppose personal identity.

3.3. Fatal error or self-identification without prior personal identity.

As Tulving, 1985 has proposed, auto-noetic consciousness is attached to mental time travel. A memory comes to mind with auto-noetic consciousness, namely with the consciousness that the subject who remembers also is the subject of her memory. But the question is: can auto-noetic consciousness constitute personal identity, or does it presuppose it? Can a subject be auto-noetically conscious of a past action she did not, in fact, do? In the Essay, II, xxvii, 13, Locke envisions the possibility of what he calls a "fatal error"¹⁷:

"But that which we call same consciousness, not being the same individual act, why one intellectual substance may not have represented to it, as done by it self, what it never did, and was perhaps done by some other agent, why I say such a representation may not possibly be without reality of matter of fact, as well as several representations in dreams are, which yet,

¹⁶ On the relation between Locke's theory of personal identity and the legal sphere, see Gordon-Roth 2019.

¹⁷ For an analysis and a discussion of « fatal error » in Locke, see for example Flew 1951; Garrett 2003; Helm 1979; Mackie 1976.

whilst dreaming, we take for true, will be difficult to conclude from the nature of things. And that it never is so, will by us, till we have a clearer views of the nature of thinking substances, be best resolved into the goodness of God, who as far as the happiness or misery of any sensible creatures is concerned in it, will not by a fatal error of theirs transfer from one to another, that consciousness, which draws reward or punishment with it.”

3.3.1. Transfer of consciousness?

This passage has given rise to a lively debate in Lockean studies and is at the heart of an important disagreement among Locke’s scholars. It is often taken as a passage in which Locke contradicts himself. According to Flew, 1951 for instance, it is a passage where “in his desperation Locke falls on his knees”. Similarly, according to Mackie, 1976, the passage on fatal error reveals the inconsistency of Locke’s view. To those stringent attacks, some Lockeans have responded. It has been proposed by Weinberg, 2012 and Garrett, 2003, for instance, that this was unfair. False memories are the result of a transfer of consciousness and thus should not count as memories.

3.3.2. First person perspective

According to Garrett, 2003, memory involves a causal condition, which is not fulfilled by false memories, and memory in Locke’s sense is not merely phenomenal memory. Again, I think that although this view is interesting, it undermines a very important aspect of Locke’s theory: I am, as a person, who I identify with, who I judge to be me (Newman 2015). Even though the identification is erroneous from the point of view of the human individual attached to it, as soon as I identify with someone, I am this person for myself. And since personal identity depends on the means of its recognition, recognizing myself in an earlier subject through a genuine or false memory produces psychological continuity with this subject of experience. I am this person as soon as I identify with her. More, as proposed by Newman 2015, who defends an epistemic interpretation of Locke’s

theory, the fact that Locke allows for the possibility of misremembering is a further reason to think that personal identity depends in Locke on an epistemic, first-person self-access. The interpretation I favor is focused on a first-person perspective, consistent with Locke's commitment to empiricism. Knowledge depends on first-person experience and perspective.

In the fatal error passage, Locke envisions the possibility of false memories, and in his view, only God can resolve these kinds of errors of misidentification. In *E, II xxvii, 24*, he writes:

“But in the Great Day, wherein the secrets of all hearts shall be laid open, it may be reasonable to think no one shall be made to answer for what he knows nothing of, but shall receive his doom, his conscience accusing or excusing him » E, II xxvii, 24.

Newman 2015 has used this passage to argue that we should favor an epistemic interpretation of Locke's theory, metaphysical facts are not within our reach, only God knows about the true metaphysical personal identities.

My understanding is slightly different: I take that here Locke does not say that God will decide, from a third person perspective who is responsible for what, and which memories are genuine and false, but only that once the secrets of all hearts revealed on the judgement day (the action is thus epistemic and not metaphysical), one's own consciousness will accuse or excuse oneself. It means that even in the judgement day, for Locke, personal identity will depend on a first-person access, made possible by a kind of Platonician anamnesia. Even in this instance, it will have to pass through the subject's consciousness.

Thus, it seems to me that Locke does not defend only an epistemic claim about personal identity but does defend also a metaphysical claim. Recognizing myself constitutes my personal identity. More, he answers to the two questions together, and overcomes the dichotomy between the metaphysical and the epistemic aspect of the question.

3.3.3. False memory phenomenology

Locke already acknowledges that it is likely that a false memory has the same phenomenology as a genuine memory. This fact is corroborated by recent literature on false-memories (Laney and Loftus 2013; Loftus and Pickrell 1995). False-memories are most of the time indistinguishable from so-called genuine memories (they are also qualified as memories by folk psychology, on this topic, see a study by Dranseika 2020, and they might even have epistemic benefits (for a discussion, see Fernández 2016; McCarroll 2017). But what we learn from the literature on false memories, is that they are not science fiction cases involving transfers of consciousness, but common cases, related to social interactions.

Consequently, if I mistakenly remember doing x or y (what Locke calls fatal error), I will self-identify with the subject of my memory, mentally place myself in the past, and have an auto-noetic consciousness of this episode, regardless of its accuracy. Then, as Loftus 1988, 2005, and others have shown, I will take this episode as being constitutive of my personal history, and it will indeed become as such. Thus, in a case of false memory, self-identification and auto-noetic consciousness do not presuppose the identity of the self but do constitute it epistemically. In virtue of the reduction claim, auto-noetic consciousness also constitutes personal identity metaphysically.

3.3.4. Objections

One could object the following: in such a case, there is not a sameness of consciousness between the remembered's consciousness and the rememberer's consciousness. Or in other words, in a false-memory case, what I discover is not myself or my personal identity, since there is no identity. To this I see one straightforward answer: there is no a priori identity, but the fact that I self-identify, even by mistake, with a subject in a past episode can make this episode constitutive of my personal past, and then ground it in its detail.

One might also object that the previous awareness condition plays an important role in Locke's understanding of memory, and that the reading I propose is incompatible

with it. If one can self-identify with something one has not been aware of in the past, then, memory cannot ground personal identity, or the account would be really weak. One way to answer such a worry is to reject the previous awareness condition (PAC): the passage on fatal error shows precisely that PAC is not a necessary condition for one to self-identify with a remembered subject of experience. PAC might be necessary for one to self-identify accurately, but it is not a necessary condition for one to self-identify. Self-recognition thus leaves room for error, precisely because it does not presuppose personal identity but results in it. In a sense, the epistemic access I have to myself can be mistaken in the sense that I can appropriate an individual act that I did not in fact do, though, once I self-identify with a remembered subject of past experience, my personal identity embraces such episodes: I feel concern and responsibility for it.

Eventually, according to the causal theory of memory, this kind of mental state might not count as memories. If a memory comes with auto-noetic consciousness, this is the consequence of the fact that this memory is caused by the very past experience of it. My memory of me doing an action *x* or *y*, is caused by my past experience of me doing *x* or *y*. There is a lively debate on the causal theory of memory, both within¹⁸ and outside of¹⁹ Lockean studies, but it is not essential here. For the purpose of this chapter, it is enough to notice that some mental episodes like false memories involve auto-noetic consciousness in the absence of a prior personal identity between the rememberer and the remembered subject. They might not be memories; they would nonetheless involve auto-noetic consciousness like other kinds of mental time travel – such as future-oriented mental time travel – do (on future-oriented mental time travel, see for instance Michaelian 2016c).

¹⁸ Don Garrett, 2003 has proposed a discussion of scholars' views on Locke's concept of memory. According to him, « memory in Locke's sense requires the revival of a perception "*that oneself had in the past*" (Don Garrett, 2003, p. 18). For an exposition of the causal view of memory in Locke see also Perry 1975.

¹⁹ For a discussion see Bernecker, 2008; Robins, 2016b; Perrin & Michaelian, 2017

Section 4 – Concluding remarks.

In this chapter, I have shown that Locke has defended three claims regarding personal identity and defined personal identity by the epistemic relation one has with herself. I have proposed that Locke has opened a path in the reflection on personal identity by focusing on the epistemic relation one has with herself as the key feature of personal identity and as the producer of this personal identity. I have shown that this epistemic relation was dependent on mental capacities such as consciousness and memory, and that Locke was reducing personal identity from a metaphysical perspective to the means of its apprehension. I have argued that the circularity objection, first addressed by Butler, 1736, was directed toward the reduction claim. Against it, I have proposed the case of false memories, where auto-noetic consciousness happens in the absence of a prior personal identity between the rememberer and the remembered.

In the cases where I mistakenly remember myself as the subject of an action or thought that I (as a human individual) did not do or think, personal identity, is, by this self-identification, recognized, and constituted. But what memory exactly is according to Locke deserves more attention. When Locke analyses the drunkard case, he sketches another feature of personal identity, mentioning witness testimony regarding the drunkard's actions. This is something that Leibniz, 1765, emphasized in his answer to Locke – others can supplement my memory when my memory is lost or misleading. The relation between personal memory and witness testimony in a psychological theory of personal identity is something that should be examined more carefully.

Another remaining question is: what happens to Locke's theory of personal identity in light of recent empirical findings about memory? More specifically, if episodic memory is not merely a reproductive capacity (as Locke and Butler thought it to be) but also has a constructive dimension, what happens with Locke's theory? On the one hand, the constructive dimension of episodic memory could strengthen Berkeley's and Reid's objection on memory's failing transitivity. On the other hand, it could be in line with Locke's view as I have described it in this paper, in the sense that memory could play two complementary roles in the relation one has with herself: episodic memory could allow self-recognition by the construction of a

representation of one's own self. Then, subsequent problems would arise, such as for instance: to what extent are persons right about themselves? Under which conditions? What is the difference between self-delusions and non-pathological constructed personal memories? This is in need of further inquiry.

Chapter 2

– Against the preservative view of memory

Introduction

In Chapter 1, I have proposed an answer to the circularity objection and started to defend a “cognitionist” (Hamou 2014) reading of Locke’s theory of personal identity. I have argued that personal identity depends on the means of its own recognition, and that memory plays a crucial role for self-extension to the personal past. But the question of what memory exactly is in Locke, and of what role it can play in personal identity is not yet answered. The aim of this chapter is to examine those two questions.

As shown earlier, according to the standard view on Locke’s, 1694 theory of personal identity, memory as the capacity to recollect past personal episodes grounds personal identity epistemically and metaphysically. I can recognize myself thanks to memory, and, as being the same person depends on self-recognition, the criterion of personal identity is memory. The attribution of this claim to John Locke

is now subject to controversies (Atherton 1983; Hamou 2014; Weinberg 2011, 2012), although even in non-memory interpretations of Locke's text, where memory is said not to play a constitutive role, it is generally recognized to play an important epistemic role. Locke's theory has received further objections (Berkeley 1732; Reid 1785): grounding personal identity in memory leads to a theoretical problem of transitivity. Reid's version, known as the brave officer objection, is the most famous and is as follows: suppose a brave officer to have been flogged at school when he was a boy, for robbing an orchard, to have taken a flag from the enemy in his first campaign when he was a soldier, and to have been made a general later. Suppose also that, when he took the standard, he had the memory of having himself been flogged at school, and that when made a general he had the memory of him taking the standard but had absolutely lost the memory of his flogging. According to Locke, he is the same person as far as his self-consciousness can extend. But then, it follows that when he was a soldier, he was the same person as the boy, and when general, he also was the same person as the soldier. However, when general, he cannot be said to be the same person as the boy, since he does not remember this past experience. So that he both is and is not the same person as the boy who was flogged. Locke's theory thus suffers from a transitivity problem.

A second objection made by Reid is against Locke's preservative view of memory: a memory cannot be the same idea as one previously formed, since according to Locke, ideas cease to be (if they are anything, they are not ideas but memory traces²⁰) before their retrieval, and according to Reid, one thing cannot have two beginnings of existence.

To overcome the difficulty of the former objection, one of the strategies is to undermine the role of memory in Locke's theory. The other one is to adapt the concept of memory. But even Parfit's and Shoemaker's quasi memory (Parfit 1984; Shoemaker 1970), or Quinton's ancestral memories (Quinton 1962) are preserved (even though not necessarily actually accessible) memory traces. Parfit's and

²⁰This is a tough point in Locke since he seems to have two contradictory intuitions: he says that ideas cease to be when they are not actually perceived, but he also suggests that memories are stored. On this question, see section 3.

Shoemaker's quasi-memories ground psychological continuity in virtue of their appropriate causal link to a past experience. They try to overcome the difficulties of the circularity objection by removing the personal part of the definition: for a quasi-memory to ground psychological continuity, it is unnecessary that there be a prior personal identity between the rememberer and the subject remembered (see Chapter 1). However, quasi-memories are still supposed to be accurate images of a past experience. Quinton's ancestral memories make room for forgetting in a memory account of personal identity. Having a memory of a past experience puts people in relation not only to this very past experience but also to the remembered subject, and her relation through memories to anterior past experiences by ancestrality. Thus, memories allow people to be in relation to themselves beyond their actual memories. They are also in relation to what they have once remembered by a relation of ancestrality.

Reid's objection against the preservative view of memory, is in fact rarely considered in the literature. Nonetheless, the current (though not recent) controversy both in philosophy and in psychology on the role of episodic memory leads one to consider it again. Against the preservative view, it has been suggested that episodic memory has a constructive dimension (Addis et al. 2009; Bartlett 1932; Daniel L. Schacter and Addis 2007; Daniel L. Schacter and Addis 2007). The ultimate aim of this chapter is to confront Locke's theory of personal identity with the debate on memory, and to propose to rethink the concept of memory. I add further arguments to Reid's criticism of the preservative view of memory. Against Locke, and relying on empirical literature, I contend that episodic memory is not a mere preservative capacity but has a constructive dimension. Nonetheless, I argue that Reid's conclusions on the nature of memory are wrong. Episodic memory does involve preservation and storage, but this does not imply that memory is only preservative. I propose that we should endorse a hybrid view of episodic memory involving both preservation and content reconstruction.

I start with an analysis of Locke's concept of memory as a repository or a storehouse (1). I argue that this leads to a content preservation problem, and I show that the objection is consistently directed toward this concept as the grounding feature of our personal identities and give an exposition of Reid's and Berkeley's objection

(2). I describe the two main routes that Locke's scholars have taken to overcome the difficulties raised by Reid and Berkeley. I argue that both face theoretical difficulties (3). I propose that the preservative view of memory not only leads to theoretical problems but also is empirically unlikely. In line with Reid's criticism, I claim that one of the main problems of Locke's theory of personal identity lies in his definition of memory. Nonetheless, I give arguments against Reid's direct realist view on memory. I propose that we should endorse a hybrid view on the nature of memory that involves preservation and construction (4) and suggest that such a view will be better equipped to solve Reid's objections (5).

Section 1 – Locke and memory as a storehouse

1.1. Memory relation of identity

Locke's theory of personal identity is widely interpreted as a memory theory of personal identity. Personal memories, as accurate representations of one's past, would allow one to access one's diachronic identity and to be an extended self. But what does personal identity depend on? As I proposed in Chapter 1, Locke's answer is this: the extension of the same consciousness, that is, memory for past actions and thoughts.

Relation of Identity: A is the same person as B, iff A truly remembers B's episodes of life as her own ²¹.

Memory, for Locke, plays a double role: first, an epistemic role: I can recognize myself over time despite the changes that affect me thanks to memory; and a metaphysical role: the criterion for personal identity is self-extension and self-recognition through same-consciousness and thus through memory. The question I am addressing in this section is, what does Locke mean by remembering?

²¹This claim has been accused of circularity (Butler, 1736), for a discussion see Chapter 1.

1.2. Remembering and the preservative view

Memory, in Locke's view, is a faculty of preservation of ideas and thoughts received by perception. In Chapter II, x of the Essay, entitled "Of retention"²², he distinguishes between two retention modes: contemplation and memory. Contemplation is a short-term memory:

"The next faculty of the mind, whereby it makes a further progress towards knowledge, is that which I call retention; or the keeping of those simple ideas which from sensation or reflection it hath received. This is done two ways. First, by keeping the idea which is brought into it, for some time actually in view, which is called contemplation." E, II, x, 1.

Locke calls contemplation the faculty of keeping a thought in mind for a while. Contemplation resembles what is called today short-term memory for information retention (Ebbinghaus 1885; Eustache and Desgranges 2015; Schacter 2008). For instance, if I want to keep in mind a phone number or a door code until I can write it down, I will use this kind of retention mode.

But the kind of retention mode that is important to his theory of personal identity is a distinct capacity: the capacity for remembering, namely the capacity to recollect personal memories – or memories of episodes that I experienced in the past.

"The other way of retention is, the power to revive again in our minds those ideas which, after imprinting, have disappeared, or have been as it were laid aside out of sight." E, II, x, 2

According to Locke, memory is supposed to preserve ideas we once had, in order to be able to retrieve them at another time. According to Locke's preservative view

²² The context of this chapter is the context of an inquiry by Locke on the faculties of human understanding. E, II, ix is devoted to the faculty of perception, E, II, xi is focused on the faculty of discernment.

of memory, for someone (A) to remember there are thus several necessary conditions. The following necessary conditions for remembering are not only Lockean but are also present (or hardly modified) in other contemporary accounts of remembering (with the exception of simulationist accounts, where at least the witness condition and the causal condition are abandoned (De Brigard 2014; Michaelian 2016a). In this regard, the issues for Locke's preservative view of memory that I present in this chapter might also apply to other accounts of remembering such as Bernecker 2009.

First, to be remembering, one must have a memory, i.e. a consciousness of a past action, namely a representation of the past. In E, II, xxvii, 13 Locke gives the following definition of consciousness of past actions (in the context of a discussion on the identity of substance): "A present representation of a past action". So, here is the first condition for remembering:

(1) *The representation condition*²³: A represents a past event (idea, action, or object) E.

Second, one must have previously been aware of E (in Shoemaker, 1970's vocabulary), or witnessed E (in (Perry, 2008)'s words), namely, one must have had the conscious experience of E. Locke writes:

"the mind has a power in many cases to revive perceptions which it has once had" E, II, x, 2

This entails,

(2) *The Previous Awareness Condition*²⁴: A must have had the idea of E²⁵.

Further, memory recollections are said to come with the consciousness that they have been experienced in the past:

²³I borrow this from Perry 2008, p. 145.

²⁴I borrow this formulation from Shoemaker, 1970.

²⁵In Copenhaver 2018's words, the Previous Awareness Condition entails that "one has an episodic memory of an event only if one was agent or witness to the event remembered".

“with this additional perception annexed to them that it has had them before.” E, II, x, 2

The present recollection of a past episode is accompanied with the same consciousness now as when the episode happened. So, memory is also a capacity to revive past episodes. It is not only a content, it is also a ‘power of reviviscence’, in Locke’s vocabulary, of this content. Memories are mental contents, but also have a specific mode of presentation: they involve auto-noetic consciousness (see Chapter 1). This term, first proposed by Tulving, 1985, is the form of consciousness that accompanies episodic memory or memory for past personal episodes²⁶, namely the consciousness that I am mentally traveling in time, that the subject of my memory is also the subject of my present thought. Memory is thus both a sum of contents related to the past, and a specific mental capacity to retrieve representations of the past with a specific mode of presentation. This leads to a further condition for remembering:

(3) *The auto-noetic consciousness condition*²⁷: A is aware that A’s representation of E in (1) is the representation of an event that A previously perceived or witnessed in (2).

Further, Locke specifies the content of a memory: it is supposed to be the retrieval of the same idea as the one previously formed. Interestingly, although Locke considers two issues of memory, such as oblivion and slowness (E, II, x, 8), he does not mention the possibility of memory inaccuracy. According to Locke, fatal errors or false memories do not fulfill the conditions of remembering. He says that they must involve a transfer of consciousness and do not count as memories (see E, II, xxvii, 13; and for a discussion see Weinberg, 2012; and Chapter 1, section X). In

²⁶I here use the most common definition of episodic memory. Nonetheless, there is a current discussion on what episodic memory is, and Tulving himself has given at least three different meanings throughout his works. First as a capacity for What where when information (Tulving 1972). Then as a capacity allowing one to revive a past experience (Tulving, 1985). And as a capacity related to autobiographical memory (Tulving 2002). For a discussion see Felipe De Brigard, forthcoming.

²⁷This is also called the retained acquaintance condition (Hoerl and McCormack 2001; Martin 2001).

fact, according to Locke, for one to remember a past idea, one must retrieve the same idea:

*“This further is to be observed, concerning ideas lodged in the memory, and upon occasion revived by the mind, that **they are not only** (as the word revive imports) **none of them new ones**, but also that **the mind takes notice of them as of a former impression**, and renews its acquaintance with them, as with ideas it had known before.”* *E, II, x, 7*

According to Locke, then, it is not only the case that remembering E gives the impression that I have had an idea of E before, it is also that no memory ideas are new ones – and this is the heart of Locke’s preservative view of memory. Memory preserves both quantitatively (sameness condition) and qualitatively (accuracy condition) a previously formed idea. So, remembering entails two other conditions: the sameness condition, and the accuracy condition:

(4) *The sameness condition*: A’s representation of E in (1) is the representation of A’s preserved perception of E in (2). The representation of E in (1) and E in (2) must be the same *token* idea²⁸.

(5) *The accuracy condition*: A’s representation of E in (1) is an accurate representation of what A has perceived or witnessed in (2).

The sameness condition is more stringent; it describes the fact that the idea perceived, and the memory of the idea are one and the same. The accuracy condition is looser; it leaves room for the fact that a memory is not exactly the same idea as an idea of perception, and even in some cases, leaves room for a small amount of memory construction (Michaelian 2016c). But the accuracy condition describes the fact that a memory preserves qualitatively the previously formed idea. The former

²⁸According to Reid, this condition entails that Locke is unable to make a distinction between the ideas of perception and the ideas of memory. In this regard, not only must the ideas in (1) and (2) be the same *token* idea, but they will end up being the same *type* of idea.

condition is endorsed by contemporary transmissionism, and the latter by contemporary preservationism (for a discussion see Michaelian and Robins 2018).

Further, Locke chooses specific metaphors to explain how a representation of E can be identical to a past idea of E: memory is a repository or a storehouse of ideas.

“This is memory, which is as it were the storehouse of our ideas.”

E, II, x, 2.

Locke’s view of memory involves content storage, in a repository or a storehouse. The metaphors echo Plato’s metaphor of memory as a pigeon-house (*Theaetetus*) and a whole tradition in which memory is described with storage metaphors²⁹. Second, they follow an old tradition in which memory is supposed to store an imprinted memory trace. Again, in Plato’s *Theaetetus* for instance, memory is first compared to a wax tablet. However, Plato’s metaphor could seem more permissive: the pigeon-house is a revision of the metaphor of the wax tablet, that allows that some memories are sometimes accessible and sometimes inaccessible, plus the idea that errors can occur at encoding, and nonetheless be stored³⁰. In Locke, memories are supposed to be accurate representations of the past, and this is made possible by the storehouse explanation of memory. If memory is a storehouse or a repository, it seems to be in a physiological sense. The repository seems to be the brain, which stores memory traces, and gives the power to revive perceptions. In Weinberg, 2012’s words, according to Locke,

“What is laid up or stored away “in memory” are traces in the brain that with an effort of the mind to revive become ideas perceived in the mind. (...) That the precursors of revived ideas are likely located in the brain (or perhaps in the “animal spirits”) is acknowledged by Locke in his admission that brain

²⁹For a history of the metaphors of memory in philosophy, see the impressive book by Douwe Draaisma, 2000.

³⁰For a contemporary reflection on the possibility of an inaccurate or constructed encoding, see the works by Chris McCarroll on the constructive encoding memory theory (McCarroll 2018).

deterioration or other diseases of the body can result in the erasure of memory.” (Weinberg 2012b:410–11)³¹.

Thus, remembering in Locke entails the following condition:

(6) *The storage condition*: E must have been stored [in A’s brain³²] when perceived under the form of a memory trace³³.

The storage condition secures the sameness and the accuracy conditions. For someone to represent an object consciously witnessed in the past, in the absence of the object, the representation of this object has to be stored somewhere. In most theories of memory, remembering is explained according to its mechanism, and also entails a causal condition (see for example Anscombe 1981; Bernecker 2008, 2009; Goldman 1967; Martin and Deutscher 1966; Parfit 1984; Shoemaker 1970; Wiggins 1967). For someone to remember a past episode, one must have witnessed it, have stored a representation, and there must be an appropriate causal relation³⁴ between the fact that A had a past experience E and the fact that A remembers E. Even though Locke did not defend it explicitly he is often said to endorse it (see for instance, Craver 2012; Garrett 2003; Weinberg 2012).

Moreover, the causal condition is integrated by neo-Lockeans in their accounts of personal identity or psychological continuity. Parfit, 1984, for example, in his adapted theory of psychological continuity, proposes that psychological continuity holds iff someone has a quasi-memory of E, someone had an experience E, and the quasi-memory (not necessarily personal) of E is caused by the fact that someone

³¹See E, II, x, 5.

³² Locke remains evasive on storage. He considers that it is an object of speculation and not of knowledge (which was true in his time). In E, I, I, 2 he writes: « I shall not meddle with the Physical consideration of the Mind”.

³³There is a debate on the format of the memory trace, whether it is picture-like or map-like; for a discussion, see Tanaka and McHugh 2018. In philosophy of memory some argue that memories are stored representations, in an picture-like format (Martin and Deutscher 1966).

³⁴For a discussion on what causalism is, see Saad forthcoming who proposes a looser version of causalism. He argues that causalism holds that “causal relations themselves” at least *partly* cause the experience.

had an experience E. The causal condition ensures the distinction between remembering and seeming to remember: the past experience must be operative in the production of a genuine memory. For one to *remember*, one's memory must be caused by one's past experience:

(7) *The causal condition*: A's memory of E is caused by A's perception of E, thanks to the storage of the impression, which secures the sameness condition and the accuracy condition.

One of the aims of the causal condition is thus to secure the distinction between cases when one genuinely remembers and cases when one only believes she remembers or one only apparently remembers but in fact makes a source monitoring error. Martin & Deutscher, 1966, have proposed a causal theory of memory, to distinguish such cases.

Let's assume for example that I have forgotten a past experience of say, freaking out when I saw an angry cat trying to get into my apartment when I was 18. Let's assume that my sister witnessed it and that later on, I forgot this experience. When I am 21, she tells me the story, I do not remember it, but I believe her. When I am 30, imagine that I have also forgotten that my sister told me the story, and that I believe I just remember it. According to Martin and Deutscher, 1966, I cannot be said to remember the experience, since the source of my memory is not the fact that I lived this experience but my sister's story. In this case, I make a source monitoring error. I believe that I am remembering because I am mistaking the source of my representation, I think it is my past experience, though it is in fact the story of someone else.

According to the causal theory, for a person to remember a past experience, they must have a present representation of the past, they must have had the past experience, and the personal past experience must have caused the memory. In the case above though, one could say that, even though I forgot that my sister told me the story, her story telling could have reactivated memory traces, such that the

memory trace of the past experience even though unavailable at one point, was reactivated and operative in producing the later retrievals. Secondly, one could argue that a memory can be caused by multiple appropriate causes, and my sister's storytelling is an appropriate cause as it allows me to form an accurate image of the past (Michaelian 2011). The causal condition is one of the most discussed conditions of remembering in contemporary philosophy of memory (for a defense of the causal theory of memory, see Bernecker 2008, 2009; for a discussion see Robins 2016b; for a critique see Michaelian 2016c), but was largely accepted until recently. It nonetheless allows, along with the storage condition, for memory accuracy³⁵, and presupposes a memory trace (Robins, 2016b).

Put together, the necessary conditions for remembering ensure that remembering is the recollection of accurate representations of the past, and thus can, for Locke, constitute personal identity epistemically. I can know myself through memory since memory is a preservative capacity that presents me with accurate pictures of personal past events.

As Locke has reduced the metaphysical question to the epistemic question (see chapter 1 on that topic), one classic objection to his memory theory of personal identity follows. The problem with Locke's theory is that it suggests that my personal identity covers all but nothing more than what I can reach by extended self-consciousness. And, if the fact I remember going to the movies yesterday generally is a justified ground for thinking that I am the person who went to the movies yesterday, what if I don't remember going to the movies yesterday: does this necessarily mean that I did not go? In other words, the question is to determine how self-access can ground the self metaphysically, since first, one cannot access everything about oneself, and second, one could be wrong about oneself.

³⁵For a non-causal account of memory accuracy, see Michaelian's reliabilism (Michaelian, 2016a), accuracy is not produced by a preservative system, but by a reliable simulational system. This view is both interesting and challenging but is the matter for a large debate into which I cannot go here, for the sake of the argument.

Section 2 – Berkeley’s and Reid’s objections

An objection of this kind has been proposed by Berkeley and Reid. In this section, I focus on Reid’s. While I believe that Reid’s direct realist view on memory is disputable, I also want to acknowledge that the objections against the preservative view of memory are strong and need our attention. I thus subscribe to Atherton’s thought, according to which “by far the most serious complaints about Locke’s theory [of personal identity] have centered around the role of memory” (Atherton 1983:275) and agree with Copenhaver 2006, 2018. She has argued that Reid’s objections are before anything else, objections directed against Locke’s storage view of memory. In this section, I show how these objections are directed against the preservative view of memory. The ultimate aim of this chapter is to give empirical arguments in support of these objections (section 4).

2.1. Reid’s exemplification of the transitivity problem.

Reid’s criticism of Locke’s theory is a critique of his use of memory in the definition of personal identity, and is twofold: on the one hand, it is related to the problem of transitivity, and to Berkeley’s objection to Locke³⁶; on another hand, it

³⁶While Berkeley’s 1732 objection is less cited than Reid’s in the literature; it nonetheless is the first one of the kind. It is thus: “Let us then suppose that a person hath ideas and is conscious during a certain space of time, which we will divide into three equal parts, whereof the later terms are marked by the letters A, B, C. In the first part of time, the person gets a certain number of ideas, which are retained in A: during the second part of time, he retains one half of his old ideas, and loseth the other half, in place of which he acquires as many new ones: so that in B his ideas are half old and half new. And in the third part, we suppose him to lose the remainder of the ideas acquired in the first, and to get new ones in their stead, which are retained in C, together with those acquired in the second part of time. Is this a possible fair supposition? You shall judge; but thus it seems to me. The persons in A and B are the same, being conscious of common ideas by supposition. The person in B is (for the same reason) one and the same with the person in C. Therefore, the person in A is the same with the person in C, by that undoubted axiom, *Quae conveniunt uni tertio conveniunt inter se*. But the person in C hath no idea in common with the person in A. Therefore personal identity doth not consist in consciousness.” (Berkeley 1732) (VII.8, 299). Berkeley’s objection is thus directed against a preservative memory theory of

is related to the preservative paradigm. The former is exemplified by the case of the brave officer:

“There is another consequence of this doctrine which follows no less necessarily, though Mr. Locke probably did not see it. It is that a man may be, and at the same time not be, the person that did a particular action.

Suppose a brave officer to have been flogged when a boy at school, for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general in advanced life: Suppose also, which must be admitted to be possible, that when he took the standard, he was conscious of his having been flogged at school, and that when made a general he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging.

These things being supposed, it follows, from Mr. Locke's doctrine, that he who was flogged at school is the same person who took the standard, and that he who took the standard is the same person who was made a general. Whence it follows., that the general is the same person with him who was flogged at school. But the general's consciousness does not reach so far as his flogging--therefore, according to Mr. Locke's doctrine, he is not the person who was flogged. Therefore, the general is, and at the same time is not the same person with him who was flogged at school.”³⁷(Reid 1785)

So, imagine with Reid that:

personal identity (for a full defense of this claim, see Gordon-Roth 2019). Such a theory leads to a paradoxical case: under certain circumstances, when forgetfulness follows remembering, two occurrences of persons do and do not belong at the same time to the same person.

³⁷Reid, in volume 1 in the seventh edition of Sir William Hamilton's *The Works of Thomas Reid, D.D.* (Edinburgh: Maclachlan and Stewart, 1872), 351

- (A) At 10 years, a boy is flogged.
- (B) At twenty he takes a standard from the enemy.
- (C) At sixty he is a general.

And then, on the one hand that in (B) he has the memory of (A). In other words, when he was twenty, he remembered being beaten when he was only ten years old. Imagine, on the other hand, that in (C) he remembers (B) but does not remember (A). So, at the age of sixty he has the memory of having taken the standard from the enemy when he was twenty, but no longer remembers he was beaten when he was ten years old. If the person in (B) is the same person as the person in (A), and in (C) the same person as the person in (B), we should say by transitivity that the person in (A) also is the same person as the person in (C). The officer being the same person at 60 and 20, and the same person at 20 and 10, he should be the same person at 60 and 10. But Locke's theory does not allow this. In Copenhaver, 2018's words: "the Memory Theory is committed to mutually incompatible theses: that the General is identical with the boy and is not". Even if memory might provide subjective evidence of personal identity, it cannot be the ground for the metaphysical definition of personal identity. Persons are not endowed with the ability of total recall. If a theory relies on a memory, defined as a collection of past experiences, representing these very past experiences and nothing more, it is a contradictory theory.

2.2. Against the preservative view.

Reid's criticism of Locke's theory goes farther: it is not only a criticism of Locke's theory of personal identity, it also is a criticism of Locke's views on memory. According to Locke's preservative view, when I remember, I revive an idea of perception that was imprinted in the brain during perception and remained available for recall.

Reid's metaphysical background is a defense of a direct realism. His views on memory follow his general ontological views: memory is not the recollection of past experiences but is a direct relation to events experienced in the past. For him,

the object of memory cannot be intra-mental³⁸, and memory is immediate³⁹. I will discuss his conclusion on the nature of memory in section 4⁴⁰, here I will focus on his criticism of the preservative view of memory.

First, according to Copenhaver, 2018, Reid's criticism is directed towards Locke's store-house view: the storage model of memory is misleading because there is insufficient evidence for positing the existence of storage of impressions in the brain. Second, even granting that they could exist, they would not be sufficient to explain memory, as they could only be correlates, and not causes⁴¹. But Reid's criticism of Locke's theory is also directed towards the sameness condition, and thus towards Locke's general preservative view of memory. He writes:

“When a thing is once annihilated, the same thing cannot be again produced, though another thing similar to it may.” Reid, 1785, p 219-220

Reid's point is thus that memories cannot be the reviviscence of the same ideas, since one thing cannot have more than one beginning. One might question Reid's background position on origin essentialism, though, and object that if ideas are individuated by content for instance, then two ideas can arise at different times but can be the same idea. Similarly, one could consider that memories are individuated by content such that a same memory could arise at different times. But those objections would focus on the content, and on the qualitative aspect of ideas and memories. And Reid's objection is directed specifically towards the sameness condition for remembering: Reid is objecting to the assumption that the idea constituting memory is quantitatively identical to the idea constituting perception.

³⁸For a discussion on the difficulties of Reid's views on memory and perception see (Folescu 2018).

³⁹For a discussion, see Debus 2008 and Sant'Anna forthcoming.

⁴⁰I do think that memories present us with representations, although Reid thinks that we remember events experienced in the past, and not ideas of events. On this matter, see again Copenhaver, 2018.

⁴¹According to Copenhaver, 2018, this part of the critique is built on Newton's rules of philosophizing: first, a theory should not posit merely theoretical causes, and second, should not posit causes insufficient to explain a phenomenon.

In this regard, according to Reid, Locke neither differentiates between perception and memory, nor does he explain memory; instead he “darkens” the notion of memory. Even in a preservative view, where impressions remain after the perception of the object, impressions being the cause and object of perception, we would not be remembering but would be perceiving⁴². Thus, Reid not only gives a famous objection against Locke’s memory theory of personal identity, but also gives a strong objection against Locke’s views on memory, and in particular against the sameness condition.

In the rest of the chapter, I will present the two main routes Lockeans have taken to answer the former objection, directed toward the memory theory of personal identity (Section 3). While focusing on the latter objection directed towards the preservative view of memory in Section 4, I will propose a third route: the concept of memory has to be rethought to play a constitutive role in personal identity. I will argue against Reid that memory involves storage. I will argue against Locke that memory is not only preservative, and I think that a hybrid view of memory can more easily accommodate the problem of transitivity (section 5).

Section 3 – Lockean responses

In this section I present two main routes among Lockeans to escape these objections. One focuses on contents and adapts the concept of memory to allow it to remain the main ingredient of the metaphysical definition of personal identity, or psychological continuity. I will present the answers of Quinton 1962; Winkler 1991 to the objections. The other route focuses more on capacities and proposes that memory is not the grounding capacity for the metaphysical definition of personal identity, though it might be necessary for the epistemic access to myself. I will discuss the propositions of Atherton, 1983, Weinberg, 2012, and Hamou, 2014. I contend that the former group fails to acknowledge the criticisms of the preservative view of

⁴²On the question of the distinction between sensation and memory, see Martin 2015.

memory, and that the latter undermines the role of memory in Locke's theory and thus faces explanatory issues.

3.1. Adapting the concept of memory.

One famous route that scholars have taken to answer Reid's and Berkeley's objections is to adapt the concept of memory, and to consider memories as embedding previous ones.

Quinton, 1962 for instance has proposed the notion of ancestry to solve the matter: when the brave officer remembers being a soldier, there is a direct continuity between him and his previous self. Since he does not remember as a 60 yo, robbing an orchard, those phases are not directly continuous. But they are indirectly related by the fact that he has a memory of phase (B) that contained phase (A). Then, it follows that, by ancestry, (A) and (B) are indirectly continuous. Quinton distinguishes two types of relations that are established between life episodes of a same person.

1. A first relation (r_1) which binds consecutive terms by memory. In (C) the brave officer remembers (B) and in (B) the soldier remembers (A).
2. A second order relation, (r_2), which binds the first term $t(a)$ and the last term $t(z)$ of a series by ancestry. $T(a)$ and $t(z)$ are connected by the relation (r_2) "ancestor of (r_1)" if each term in the series $t(a), t(b)... t(z)$ is related to the following one by (r_1).

Even if memory is missing for the oldest events, the term-to-term relationship between all the life episodes of the same person makes it possible to establish (r_2), and a form of weak continuity that is sufficient to explain the unity of a person, and gives a basis for the judgment that one is the same person.

There are other answers of this kind. Winkler, 1991's answer to the objection of the brave officer goes in a similar fashion: Winkler, who has proposed that Locke's

theory is an appropriation theory of past actions and thoughts, suggests that when the general identifies with the young officer, he “commits himself to the actions the officer appropriated” (Winkler 1991, p. 222). In Winkler’s explanation, self-identification in a memory involves an identification not only with the remembered subject of this past event, but also with the history of the remembered subject.

Nonetheless, Forest proposed a further objection: “the senile general, and the vagaries of memory”⁴³. Memory and oblivion are not distributed systematically among events, and we do not forget only older episodes. Let us go back to Reid’s case and to the three life episodes (A), (B) and (C). Imagine here that in (C) the general remembers (A) but that he no longer remembers (B). The second-order relation cannot do the job of binding episodes, since the chain of memories is broken.

However, one could answer that the case is underspecified in a way that still allows the ancestral relation to bind all the person episodes together. Suppose there is a period of time between A and B such that the A person remembers it and the person in it remembers B, A could be bound to B by any in-between ancestral memory.

But let’s imagine a more radical case. Let’s imagine that I have an experience for which I have never formed a memory and will never remember it. Let’s just mention, that this kind of case might happen and is called selective memory. Memory encoding seems to be specifically related to my attention (see section 4), such that I do not remember everything because I do not pay attention to everything. The ancestrality relation will not do the job of binding myself with a past experience I did not pay attention to. So that a memory theory of personal identity will fail to explain the identity of persons outside of the reach of their attention and conscious experiences.

One answer could be this: attention, consciousness and memory should be thought of in terms of potentiality and not as necessarily actual. For example, I may not be remembering that I woke up brutally this morning, but that would not mean that it

⁴³ Forest, course on personal identity, University Lyon 3, 2011-2012.

was not I who woke up brutally. Indeed, I can never, and Locke had already pointed this out, have an overview of either my present or my past. That does not mean it is not mine. In *E*, II, xxvii, 9, Locke writes:

"As far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person."

E, II, xxvii, 9

Locke uses “can” in order to express potentiality. Personal identity would thus rely more on the potentiality of consciousness and remembering than on their actuality. This should lead us to modify the first statement of the personal identity relation, consistent with Locke’s view:

*Relation of Identity**: occurrences of person A and B are occurrences of one and the same person if and only if A contains in fact or potentially B’s consciousness for past events or B’s memories⁴⁴.

This could be an answer and would save memory from the brave officer and the senile general objections, but this would still fail to acknowledge Reid’s criticism against the preservative view of memory. Quinton’s and Winkler’s answers modify the concept of memory: one memory could be embedded in another one. This seems like a good start to go beyond the preservative view of memory and gives an answer to the transitivity objection. But it does not answer to the other part of the objection directed towards the preservative view of memory and the sameness condition. Such revised memory theories of personal identity still rely on the idea that memories, when they occur, are pictures of real past events.

⁴⁴A similar answer has been proposed by Grice 1941: Grice proposes the notion of “total temporary state” or t.t.s., which results in this relation in Perry’s words: “RG: There is a sequence of t.t.s.’s (not necessarily in the order they occur in time, and not excluding repetitions), the first of which is A and the last of which is B, such that each t.t.s. in the sequence either (i) contains, or would contain given certain conditions, a memory of an experience contained in the next, or (ii) contains an experience of which the next contains a memory, or would contain a memory given certain conditions.” (Perry 2008, p. 136).

3.2. De-emphasizing the role of memory: consciousness is the key ingredient of personal identity.

A second route proposes that in Locke's theory another capacity other than memory in fact makes personal identity.

Margareth Atherton, 1983, has proposed what has become an influential view in Locke studies, that in Locke, what makes personal identity is not memory but is consciousness.

Further, Shelley Weinberg develops a defense of Locke's view relying on consciousness and not on memory⁴⁵. In Weinberg, 2012's view of Locke, one's consciousness for one's momentary mental states cannot be enough to determine diachronic identity, or Locke could not give an account of just punishment and reward.

This actually gives an answer to Reid's and Berkeley's objection: "we are not presently aware of all we have done" (Weinberg 2012, p. 389). If consciousness of past and present mental states was enough, we could not justly solve two parallel problems: one is the memory loss for a crime I committed (I would unjustly escape punishment), and one is the false memory of having committed one I did not in fact commit (I would unjustly be punished).

Thus, subjective access to personal identity cannot be enough, "rectification is impossible, for there is nothing objective for God to appeal to in determining whether or not we should be punished" (Weinberg 2012, p. 389). So, according to Weinberg, there is an objective fact of an ongoing consciousness in Locke, that makes his theory coherent. Consciousness is a self-referential awareness internal to every perception and ongoing through these momentary perceptions.

⁴⁵She follows Atherton in her distinction between consciousness, thinking, and memory, though she gives arguments against the psychological view of consciousness. According to Weinberg, consciousness also is a metaphysical fact (Weinberg, 2011).

In this regard, the consciousness we access through momentary conscious states is part of this wider ongoing thing that supports all of our perceptions. We have an epistemic or a phenomenological access to what she calls the subjective fact of consciousness, that allows us to access also to a metaphysical fact of an objective ongoing consciousness:

“I am conscious I am perceiving an idea I had before. Thus, I have a unified experience of being conscious of two perceptions of ideas as opposed to an experience that would be described as including two perceptions of ideas with two consciousnesses... We have the experience of ourselves as existing diachronically – as temporally extended.” Weinberg, 2012, p. 403.

Her interpretation relies on the revelation of an ambiguity in Locke’s concept of consciousness: consciousness is both “a momentary subjective experience” and “the objective fact of an ongoing consciousness” (Weinberg 2012, p. 390). But the latter is the ingredient she needs to answer to the objections of violating transitivity and save Locke from inconsistency. If Locke’s theory is grounded on consciousness as an objective fact, my epistemic access or incapacity to access specific episodes does not change anything for my personal identity. If an episode has been experienced by my consciousness, even though I do not actually remember it, it still belongs to me, since it was supported by this same ongoing consciousness.

A closely related interpretation has been proposed by Philippe Hamou, 2014. In his view, memory does not solve the metaphysical question of personal identity but “the same train of conscious thoughts” does. According to him, although psychological continuity does ground personal identity in Locke, it does not depend on the subjective authority of actual memory. Hamou’s interpretation relies specifically on the comparison between chapter E, II, xxvii, and chapter E, II, xiv. According to Hamou, E, II, xiv specifies what Locke means by ‘consciousness’; for instance, on E, II, xiv, 4, Locke writes:

“When that succession of ideas ceases, our perception of duration ceases with it; which every one clearly experiments in himself, whilst he sleeps soundly, whether an hour or a day, a month or a year; of which duration of things, while he sleeps or thinks not, he has no perception at all, but it is quite lost to him; and the moment he begins to think again, seems to him to have no distance.” E, II, xiv, 4.

To Hamou, this passage shows what Locke means by consciousness: a train of ideas constantly succeeding one another. By comparing E, II, xxvii with E, II, xiv, Hamou proposes that in Locke the succession of ideas in chapter E, II, xiv, is the same continuous consciousness in chapter E, II, xxvii. Thus, what grounds personal identity is the objective fact of a same train of conscious thoughts. This reading de-emphasizes the role of memory in Locke’s theory of personal identity. However, according to Hamou, memory still plays a very important role: it is the means by which a consciousness can self-testify to a diachronic identity. In this regard, Hamou’s position is close to Reid’s, for whom memory does provide evidence for personal identity from a subjective point of view. But one problem for a memory theory of persons, and one reason why it could not play a constitutive role in Locke, is that self-identification in a memory could be misleading⁴⁶.

“The memory is always a fact of present consciousness, and thus, does not carry the absolute guarantee that our remembered ideas have been lived in the same train of consciousness as the one that today remembers them.”⁴⁷ (Hamou 2014:21).

Hamou refers here to the possibility of formation of false memories, where I self-identify by mistake in a memory. To him, this should be taken as evidence for the fact that Locke did not defend a memory criterion of personal identity. Because, for memory to play this role, it should be necessarily immune to errors through misidentification – or it would be unable to guarantee personal identity. Since this fact is not secured, memory cannot ground personal identity in Locke, or his theory would fail.

⁴⁶For a discussion on auto-noetic consciousness and false memory, see Chapter 1.

⁴⁷My translation.

By considering memory errors, Hamou is thus the closest to acknowledging Reid's criticism of Locke's preservative view of memory. But what Hamou acknowledges in fact, is not exactly that a memory cannot present us with an idea already perceived (against the sameness condition) or can be a partly inaccurate representation of one's past (against the accuracy condition), but that episodic memory is not immune to errors through misidentification (against the auto-noetic consciousness condition). He thus envisions the possibility of a mistaken self-identification through memory but does not answer Reid's objection to the idea that a memory content is the same as the content of a past perception. And his subsequent strategy is to de-emphasize the role of memory for the constitution of personal identity in Locke.

In fact, Atherton, Weinberg and Hamou all think that if memory plays a role in Locke's theory it is as a capacity that allows one to have an epistemic (subjective in Hamou's vocabulary) access to oneself. But it cannot ground personal identity metaphysically, namely cannot be constitutive of one's personal identity, or the theory is threatened by objections. Memory can answer the epistemic question of personal identity but cannot give an answer to the metaphysical question. Their readings rely on the idea that in Locke memory and consciousness of the past are distinct. As shown in chapter 1, this is true of some passages, but first, the distinction between memory and same consciousness is not incompatible with a memory interpretation of Locke's theory. Second, as I promised to discuss, some other passages in Locke seem to say otherwise and to equate memory and extended consciousness. In particular, in the Day and the Night-Man passage, he writes

“For granting that the thinking Substance in Man must be necessarily suppose'd immaterial, 'tis evident, that immaterial thinking thing may sometimes part with its past consciousness, and be restored to it again, as appears in the forgetfulness Men often have of their past Actions, and the Mind many times recovers the memory of a past consciousness, which it had lost for twenty Years together. Make these intervals of Memory and

Forgetfulness to take their turns regularly by Day and Night, and you have two Persons with the same immaterial Spirit, as much as in the former instance two Persons with the same Body. So that self is not determined by Identity or Diversity of Substance, which it cannot be sure of, but only by Identity of consciousness.” in E, II, xxvii, 23

Here, Locke describes the lack of memory and the lack of consciousness for the past as one and the same thing. Atherton and Weinberg themselves acknowledge that some passages, and this one in particular, trouble their readings:

According to Atherton, 1983:

“Perhaps the most troublesome passage for reading Locke’s views on memory in the way I have suggested comes at 2.27.23.” (Atherton 1983, p. 285)

And to Weinberg, 2012:

“Perhaps the most difficult passage for my interpretation is Locke’s example of the “Day and the Night-man”, in E II, xxvii, 23, where Locke seems to equate a lack of consciousness with forgetfulness. Again, I would argue that in these cases Locke is focusing on the first-personal awareness of ourselves”. Weinberg, 2012, p. 406, footnote 49.

But the thing is that in other occasions in E, II, xvii, and for example in the Nestor and Thersites and in the Prince and Cobbler passages, Locke focuses on the first-personal awareness of ourselves. In the Nestor and Thersites passage, Locke writes that should I be conscious of Nestor’s past actions in a first-person perspective, I would be Nestor.

“Let him once find himself conscious of any of the action of Nestor, he then finds himself the same person with Nestor.” E, II, xxvii, 14.

Similarly, in the Prince and the Cobbler passage, the thought experiment is thus: imagine that the Prince and the Cobbler swap bodies, such that the Prince’s consciousness is transferred into the Cobbler’s body and vice versa. According to

Locke in this case, although for witnesses the Cobbler will be, as a human, where his body is, his consciousness will determine who he is as a person. If the Cobbler in the body of the Prince has the memories of the Cobbler and can extend himself to the past by self-identifying with the Cobbler's past episodes, he is the same person with the Cobbler and not with the Prince.

Moreover, in E, II, xxvii, Locke generally seems to reduce the metaphysical definition of personal identity to the means of self-access. For instance, in E, II, xxvii, 24, he writes:

*“if there be any part of its [its is for the self here] existence, which I cannot upon **recollection** join with that present consciousness, whereby I am now my self, it is in that part of its existence no more my self than any other immaterial being. For whatsoever any substance has thought or done, which I cannot **recollect**, and by my consciousness make my own thought and action, it will no more belong to me, whether a part of me thought or did it, than if it had been thought or done by another immaterial being any where existing.” E, II, xxvii, 24.*

What makes myself extended to the past metaphysically and not only epistemically, is thus the possibility of recollection. My personal identity extends to the past as long as I can recollect my past thoughts and actions, with auto-noetic consciousness, namely with the consciousness that I was the person who had or did those thoughts and actions.

Memory in Locke seems to be a very important ingredient of personal identity. Without memory it seems unclear how a subject can access the details of her personal history, and how a subject is an extended full subject. Locke's theory is an epistemic theory of self-access (see for example Newman 2015; Schechtman 2016), where the person is not an empty subject but a subject with a personal history she can relate to. As I have argued in Chapter 1, in Locke, the answer to the epistemic question about personal identity also gives an answer to the metaphysical question. Personal identity depends on the means of self-access. Abandoning memory as a

grounding ingredient of Locke's theory of personal identity, one has difficulties in accounting for how we relate diachronically to ourselves in detail⁴⁸.

So, there are two types of strategies among Locke scholars to overcome Berkeley's and Reid's objections. One keeps memory as the grounding capacity of personal identity and adapts the concept of memory, the other one adapts the theory and argues that Locke did not defend a memory theory of personal identity. The latter faces problems in explaining passages where Locke seems to use memory and consciousness of the past to mean one and the same thing. The former fails to acknowledge the second part of Reid's criticism against the preservative view of memory. In any case both have a preservative view of memory.

Section 4 – A hybrid view of episodic memory

From now on, I intend to propose a third route to answer to Reid's objections. I argue that Reid's criticism of the memory trace is not in line with recent empirical findings. I show that according to recent empirical literature on episodic memory, there are such things as memory traces. But the memory trace does not necessarily involve accuracy, and certainly need not involve the sameness condition. The sameness condition stipulates that for someone to be remembering, the representation in (1) must be the representation of the same idea as the idea witnessed in (2). Beyond Reid's theoretical objection to this condition, the ongoing research on memory gives further evidence against it. Thus, Locke's preservative view of memory is unlikely. Let us recall the story. For episodic memory to ground personal identity and/or to be a source of self-knowledge, it has to be preservative.

⁴⁸This remark might also hold for attempts to solve the problems of personal identity by appealing to the minimal self. Even though minimal definitions of the self are appealing in many regards, since they seem to overcome objections addressed to psychological theories of personal identity (see for example Viera unpublished, it is unclear how they could explain how people access and relate to the detail of their diachronic identity as temporally extended persons. In this regard, such accounts cannot replace accounts of personal identity, since they do not exactly have the same object.

Namely, a memory has to represent either the same idea as – or in a weaker view, an accurate representation of – one’s previous perception of a past event. And according to the preservative view of memory, this is made possible by the fact that the representation of the event witnessed in the past is stored in the brain, until the memory trace of the past experience causes remembering (under specific conditions). So that the witness condition and the storage condition secure the sameness condition, or the accuracy condition⁴⁹.

In this section, I will rely on two main findings of the last century in memory science to advocate for an alternative explanation of the nature of episodic memory: episodic memory has a constructive dimension: encoding, storage and retrieval are partly constructive processes; the preservative view is thus misleading. But against the direct realist view, episodic memory also involves storage.

First (4.1), psychology of memory seems to support Reid’s objection to the preservative view of memory. Episodic memory does not seem to present us with the exact same ideas as the ones previously perceived or formed.

But, (4.2), Reid’s conclusion on the nature of memory is unlikely. What we learn from neuroscience is that episodic memory involves storage. Memory does not put us in a direct relation to past events, but in a mediated relation to past events through memory traces. Memory is thus preservative in a sense, but this is not the whole picture.

Storage is not neutral (4.3). We are not presented twice with the same idea; the sameness condition is not secured by the episodic memory system. This is not because of the absence of a memory trace but because of the nature of episodic memory. I will focus on two phenomena described in neuroscience, consolidation and semantization processes, to argue that memory storage does not secure memory accuracy. Thus, the preservative view of memory is unlikely, not entirely false

⁴⁹The criticisms I address against the preservative view in this section also apply to contemporary memory epistemic generativism (Bernecker 2009), according to which those conditions do secure the accuracy condition through a causal law according to causal theories (Bernecker 2008).

though, but unlikely because it is only one part of the story. Episodic memory also has a constructive dimension.

4.1. First, the witness condition does not secure the sameness condition

The psychology of memory seems to be in line with Reid's objection: memory does not present us with an accurate picture of what we saw, or with an accurate idea of the event we witnessed. It is not because I witnessed an event that I will remember it accurately, or that my representation of it in a memory will be the same representation as the one I had when I perceived it.

4.1.1. Eyewitness testimony and memory distortions.

Research on eyewitness testimony is interesting on that topic, and an old field. An experiment conducted by the criminologist von Liszt in 1902 is considered to be the first of the kind. During class, two students start to argue, one of them ends up drawing a gun, and a professor steps between them. Other students, witnessing the scene, are shocked. This was a rehearsed scenario. Liszt asked students to write exactly what happened, and their testimony was compared with the original script. The most accurate witnesses had an error rate of 25%, the worst 80%⁵⁰ of the details, when their reports were compared to the original script. We thus learn two things: witnessing does not entail total accuracy, and accuracy comes in degrees.

Another pioneer experiment was conducted by Cattell in March 1893. Cattell asked 56 students to answer questions such as 'What the weather was a week ago?'. To this question 16 answered 'clear', 12 said 'rainy', 7 said 'snow', 9 said 'stormy', 6 said 'cloudy', 6 said 'partly stormy partly clear'. The answers were covering all the possible March (where?) weathers. Inaccuracies in witness reports for weight, speed, time and distance have been shown to be common (Bartol and Bartol 2006),

⁵⁰For more details on that experiment, see Loftus 1980, p. 20 & sq.; Münsterberg 1908, p. 50 & sq.; Schulz 2010, p. 542 & sq.).

and even more common and more important under stress or anxiety conditions (Sarason and Stoops 1978). Witnessing a violent event also seems to increase inaccuracy reports (Clifford and Scott 1978). For a review of the psychological literature on witness reports, see (Loftus 1980, p. 33 & sq.).

Since those experiments, eyewitness testimony has been studied widely and systematically. Liszt's and Cattell's conclusions hold in more recent work on eyewitness testimony. We take our memories as a reliable source of knowledge, but we overlook the fact that episodic memory has a constructive dimension. Witnessing a scene does not ensure that we will remember it truly or accurately⁵¹.

The picture of episodic memory that emerges from the last century's research in psychology is the picture of a highly constructive capacity. Some psychologists even defend the claim according to which memories are all to some extent false (Loftus and Pickrell 1995). So that memory does not present us twice with the same idea. But one could wonder whether the research on eyewitness testimony is evidence for claiming that memory does not present us twice with the same idea. It could be, instead, evidence for the fact that encoding is constructive, i.e. perceiving involves the production of a non-totally accurate representation of an object, scene or event. It could say nothing about the constructive dimension of episodic memory; episodic memory could still be preservative. So that the reason why memory outputs are inaccurate might not be because of episodic memory construction at retrieval but because of a constructive encoding.

4.1.2. Constructive encoding

Is it memory which is a constructive process or is it perception? It could appear that memories are inaccurate because encoding was inaccurate. Or in other words, what is constructed is not what we remember but what we see. To illustrate this point,

⁵¹Research on eyewitness testimony presupposes that memory construction happens at retrieval. But McCarroll, 2018, for instance, has proposed that construction could happen either at retrieval or at encoding, namely at the moment we perceive. On the constructive dimension of perception, and on amodal completion, see Nanay 2018a, 2018b.

here is a ghost story. In *Ghosts* (Calle 2013), Sophie Calle put together collected descriptions of removed, lost, or stolen paintings in museums during the 1990's: at the Isabella Stewart Gardner museum in Boston, at the Earl of Bath's residence in England, or at the MoMA for the exhibition *Dislocations*. She has collected descriptions of the paintings from members of the staff, guards, janitors, officials, etc. About Magritte's *Menaced Assassin*, here is how four people remember it:

“– It is a painting with a smooth surface, an easy one to spot check. It is approximately five feet high and seven feet long. (...)

*– It has a film noir sort of feel, a mystery novel look to it. The puzzle is here. You have all those little clues that probably won't lead you nowhere; there are men dressed in dark coats and black bowler hats, the way Albert Finney was dressed in *Murder on the Orient Express*, placed in a room with a dead body. In the center, the one who seems to be the perpetrator is lifting the needle of a photograph. Two weird-looking individuals are hiding to the side. There is a face looking from the balcony, almost like a sun on the horizon. And, when you look at her carefully, you realize that the towel probably conceals a decapitated head.*

– It's just one more picture where the woman is naked and the men are clothed.

– I think this is just a murder scene. Men in dark suits, a pale woman and dashes of red blood. That's all I remember.”(Calle 2013)

In his response to Skinner about language, Chomsky suggested that verbal responses depend on the background of the subject, their histories, intentions, beliefs, etc., which are all factors that can influence a person's verbal response (Chomsky 1959; Skinner 1957). This suggestion seems to apply here. It seems clear that the subjects do not have the same background and that their recollection of the painting depends at least in part on what interested them when they were looking at it. Memory encoding therefore seems to depend on the history of the individuals. So that episodic memory could very well be a preservative capacity and when

memories are distorted or at least selected representations of a past event, that could be in virtue of the preservation of an already distorted or at least selected representation of the world.

Encoding has been shown to be selective, and dependent on bias. For instance, it depends on attention. Studies have shown that self-related events, and especially self-active involved events, were more likely to be remembered, because they were more likely to be encoded (Mulligan and Hornstein 2003; Seamon, Philbin, and Harrison 2006). In line with research on the constructive dimension of perception, McCarroll, 2018 (who does not deny that retrieval can also have a constructive dimension), has proposed the hypothesis of constructive encoding to disentangle encoding and retrieval construction.

4.1.3. Reconstruction in Repeated Reproduction

But another thing we learn from psychology is that several instances of episodic memories targeting a same event, image or story are unlikely to be exactly the same. So that it could appear that memory is constructive both at the level of encoding and at the level of retrieval. On this matter, a second ghost story is interesting, the one used by Bartlett. Bartlett was a British psychologist who is considered a precursor of cognitive psychology. While Bartlett has also worked on body memory and memory of gesture, his work on declarative memory and on episodic memory is unique for his time, when behavioral psychology focused on motor learning and motor responses to stimuli.

Bartlett, 1932 conducted an experiment using the "Repeated Reproduction Method", in which he had twenty of his students read *War of the Ghosts*, a traditional North American tale translated by Franz Boas, and then had them reproduce it at several later times (Bartlett 1932). The first reproduction of Subject H took place twenty hours later. Bartlett notes that his story is shorter than the original one, his style is more modern – almost "journalistic," – some elements have been omitted, and others transformed. For example, "canoe" became "boat," "seal hunting" became "fishing," "Egulac" became "Edulac," "Kalama" became "Kaloma. The second

reproduction by H took place eight days later. The distortions of the first reproduction seemed to have become more pronounced. The narrative is even shorter, but it is also more coherent. "Kaloma" has disappeared, "the arrows" have also disappeared.

Two things are specifically interesting: first, memory construction seems to depend on the subject's background. The students' stories are always modified to some extent, names are regularly changed, and transpositions into a vocabulary more familiar to English students, such as the transposition of "canoe" to "boat" or "hunting seals" to "fishing" for example, have many occurrences. Second, even if encoding is constructive, Bartlett's experience leads one to think that retrieval also is.

The description of episodic encoding and retrieval seems to plead in favor of Reid's argument against the preservative view of memory: episodic memory cannot present us twice with the same idea. But it does not necessarily plead in favor of his anti-storage conclusion. In Reid's view, memory cannot be preservative, but it is not constructive either, it is a perceptive capacity. For Reid the argument against the sameness supports his direct realist views: memory cannot present someone with a previously formed idea. Memory puts someone in a direct relation to the past.

If the argument against the sameness condition holds, it does not entail direct realism. It can entail a constructive view of episodic memory. One could propose a hybrid view: episodic memory is the retrieval of a past perception, but the act of recollection transforms the idea (representation, or mental imagery) recollected. This view is actually more consistent with recent empirical literature on episodic memory. An episodic memory cannot be the recollection of the exact same idea, either quantitatively or qualitatively, that one perceived in the past. But it is not because memory presents us directly with the past, it is because the mental action of recollection is not neutral.

As perceiving is partly constructing an image of the world, remembering is partly constructing an image of the past. Bartlett has shown that remembering is a constructive capacity in the sense of an interpretative capacity, where people recall, or form a representation of the information encoded, according to their cultural

schemas (Bartlett 1932). Schemas are generally helpful for several cognitive tasks (organizing and interpreting information for understanding, retention and recall), but produce memory distortions. A dysfunctional feature of memory regarding accuracy in details is the result of memory efficiency in other regards, and thus of an adaptive capacity.

4.2. Memory storage

Episodic encoding and recall are constructive, but this is not the whole picture. Reid's conclusions on the nature of memory are unlikely. Memory does not put people in a direct relation with past events but appears as the retrieval of encoded and stored mnemonic traces.

At least since the case of HM, for whom an ablation of the hippocampus resulted in an episodic amnesia, the relations between episodic memories and the hippocampus have been a subject of investigation. HM are the initials of a patient (Henry Gustav Molaison) suffering from severe epilepsy. In 1953, at the age of 27, he had a surgery: a lobectomy aiming at curing his epilepsy. A lobectomy involves the removal of parts of the brain: in this case, the operation consisted of removing a set of structures from the medial region of the temporal lobes on both sides of the brain including most of the hippocampus, the amygdala and some adjacent areas of the temporal cortex.

After the operation, HM sank into a deep but not total amnesia:

- (1) His motor skills and habits were intact.
 - (2) If questioned immediately after listening to a series of numbers, for example, he was able to recall them as well as non-amnesic subjects. His working memory seemed to be intact.
 - (3) A lot of his general knowledge was preserved.
 - (4) He had memories of several episodes that occurred before he was 16 years old.
 - (5) But he could not remember specific events in the decade before his operation.
- His amnesia was said to be retrograde.

(6) And his ability to form new memories was also affected; he could not form new memories for events that happened to him after the operation. His amnesia was also said to be anterograde.

Because HM disorders are quite selective (his intelligence, IQ, perceptive functions, comprehension, linguistic production, habits, and many parts of his knowledge were spared), they have been taken as evidence for at least two facts: on the one hand HM is a support for theorizing distinctions between different types of memory (procedural memory, short-term memory and long-term memory (Squire 1992)). On the other hand, from his case, amnesia appears as an effect of hippocampal lesions, and conversely, the hippocampus is thought to be responsible for encoding, retrieval (Nadel and Moscovitch 1997) and consolidation (Eustache and Desgranges 2015; Squire 1992).

In a similar trend, Tulving 1985, made a distinction between three different types of memory. Originally, Tulving based himself in particular on the symptomatic case of NN. After a car accident NN's hippocampus was damaged. He became deeply amnesiac with regard to personal past events before and after the accident. His short-term memory was impaired, and he had difficulty in fixing new memories. Moreover, even though he still had some knowledge about his past, it seemed to be pictured in the same impersonal color as his knowledge of the rest of the world. NN's injury affected his personal memory but neither his habits nor his mental encyclopedia (as it was also the case with HM). NN, HM and other people with episodic amnesia (Squire, 1992) have become evidence for the fact that episodic memory is (1) supported by a specific set of brain regions and (2) partly independent from other memory systems such as procedural and declarative memory.

Thus, Reid's conclusions on the nature of memory are unlikely. What we learn from neuroscience is that episodic memory involves storage and is supported by a specific set of brain regions. Memory does not put us in a direct relation to past events, but in a mediated relation to past events through stored memory traces.

4.3. A hybrid view of memory⁵²

Memory is thus preservative in a sense, but this is still not the whole picture. I have already shown that encoding and retrieval have a constructive dimension. Here, I intend to show two last things. First, storage is not neutral either. Second, the question remains of whether this is a sign of the weakness of memory or whether this is a sign of another function of episodic memory. What we learn from current neuroscience is that episodic memory has a non-dysfunctional constructive dimension. Memory construction is not always a weakness of memory but in certain cases is the sign of a well-functioning memory. Thus, it is unlikely that memory presents us with the exact same idea as one formed through perception, either quantitatively (ideas of perception and ideas of memory are of different types), or qualitatively (the content of those ideas is likely to be at least partly different). Neither the sameness condition nor the accuracy condition is secured by the episodic memory system. This is not because of an absence of memory trace, though, but because of the constructive nature of episodic memory.

4.3.1. Storage, consolidation and semantization

First, storage is not neutral. This is what we learn from the work on the processes of consolidation and semantization. The idea of memory consolidation, related to the idea that memories need time to become less vulnerable, is an old idea (see for instance Burnham 1904). Recently, researchers have made the distinction between synaptic and system consolidation. Synaptic consolidation refers to the quick neural mechanisms that underly memory formation: it is also called late-phase Long Term Potentiation (LTP) (Frey, Huang, and Kandel 1993; Huang and Kandel 1994; Lu, Kandel, and Hawkins 1999). LTP refers to a synaptic strengthening that produces an increase in signal transmission between neurons and is described as underlying memory at the cellular level (Craver 2003; Lømo 2003; Lynch 2004). System consolidation refers to the reorganization of memory in the brain through time and life that makes some memories less vulnerable than others.

⁵²For an alternative view see Michaelian 2016c; for an alternative hybrid view, see Sant'Anna 2018. I will discuss these views in more detail in Part 2.

Nadel and Moscovitch, 1997 have proposed the Multiple Trace Theory. According to this theory, each time I recall a specific episodic memory, a variation in its trace is formed, which is both added to the memory trace, and reinforces it. As the multiple traces are produced, corresponding to the multiple retrievals of this memory in different contexts, the memory is both consolidated and transformed. More recently, people have proposed the transformation hypotheses (Wiltgen and Silva 2007; Winocur, Moscovitch, and Bontempi 2010), according to which specific memories after several retrievals can become more abstract representations, formed from all the corresponding memory traces, that capture the gist but few of its details. Winocur and Moscovitch 2011 have proposed that some episodic memories are semantized over time, i.e. that, following successive recall, their traces progress to extra-hippocampal structures and become semantic memories (though that does not necessarily erase the episodic memory hippocampal traces; they can coexist, Winocur & Moscovitch 2011).

But one question remains: is the memory transformation and reconstruction into a more schematic or gist-based representation the sign of a weak memory system?

4.3.2. The adaptive dimension of memory distortions

Reflection on non-dysfunctional memory distortions has a long history and can be dated at least back to Ebbinghaus. Ebbinghaus inaugurated research on list-retention and showed that above a certain number of items in a list, we generally did not recall the details of the list (Ebbinghaus 1885). Nonetheless, we can recall a general sense of the list (Roediger and McDermott 1995). Schacter, 1997, following Roediger and McDermott, 1995, has proposed the following list:

“There is a good chance that you can experience memory distortion of this sort by yourself by paying careful attention to the following series of words: candy, sour, sugar, bitter, good, taste, tooth, nice, honey, soda, chocolate, heart, cake, eat, and

pie. Turn away from the page now and take a minute or so to write down all the words you can remember from this list.

Now, take the following test. Consider the three words printed in italics at the end of this sentence and, without looking back to the previous paragraph, try to remember whether they appeared on the list that I just presented: taste, point, sweet.” (Schacter 1997)

In the false recognition case, we believe that we recognize when we in fact associate. Nevertheless, should we think that is a memory failure? If we say so, then we presuppose that memory must be preservative, both quantitatively and qualitatively. But you may have believed you recognized “sweet”; however, you did not make a mistake about “point”. Thus, you got the general category, and you understood its meaning. Your memory retained it, and in this perspective, we cannot really say that your memory is deficient, or we would miss a point in its description. It seems more appropriate to say here that memory is selective because even if it does not retain everything, it does not retain whichever, nothing or nonsense. Thus, memory would not retain everything but would have complementary strategies in order to retain needed information. A dysfunctional feature of memory regarding accuracy in details is the result of memory efficiency in other regards, and thus might not be dysfunctional.

Remembering the gist seems to enhance remembering, such that memory construction processes support the ability to remember. Interestingly, amnesiac patients with hippocampus damage do fewer false recognition errors than do control subjects. They appear to have low memory for the gist. It has thus been proposed that gist-based and associative memory are indications of a healthy memory (Gutchess and Schacter 2012).

Moreover, true and false recognition results show similar levels of brain activity in the hippocampus and other regions involved in accurate remembering (Addis, Wong, and Schacter 2007). Research has shown that associative processes improve memory performance by allowing organization of information (Schacter, Guerin, and St. Jacques 2011), and that gist-based processes improve the capacity of

abstraction by allowing generalization from separate pieces of information. The idea that episodic memory has a constructive dimension and aims not only at accuracy in the details but also at consistency is thus largely accepted and has been called ‘the modern view on memory’ (Conway, Singer, and Tagini 2004). Episodic recall will never be totally accurate in the details but will be an – always to some extent – inaccurate representation of the past (Conway 2005). In this regard, as it is theoretically costly to say that episodic memory aims at accuracy but always – to some extent – fails, both at encoding and recollection, we might be forced to adopt a less ambitious claim about episodic memory functions, and consequently a less ambitious claim about the epistemic value of remembering.

Section 5 – Back to the objection

I have given reasons to endorse a hybrid view on memory. I have contended that memory is a capacity that preserves memory traces which are formed and transformed during encoding, storage, and retrieval. Consequently, the last section led to rejecting a mere preservative view of memory and put pressure on the sameness condition for remembering. The sameness condition for remembering cannot be fulfilled in virtue of the nature of memory. The last section also raised issues concerning the accuracy condition. The accuracy condition should be considered carefully: content inaccuracy concerning some aspects (for instance the detail) can result in accuracy regarding other aspects (for instance the gist).

Now, let’s go back to the case of the Brave Officer. If memory has a constructive dimension, and if the picture drawn above is right, it is likely that specific autobiographical memories of his own childhood have been semantized into the category ‘memories of childhood’, such that the general is related to the childhood episode through his general knowledge of his past, even though he cannot actually recollect this very past experience.

Second, it is likely that, as he has retrieved this specific memory previously, some of his other memories are related to this one and have integrated pieces of information coming from this source, even if indiscernible. So that, because he encoded the representation and remembered it at least once, it has become material for other mental states and representations of his past. Memory material is recombined and reconstructed to form representations of the past but might be recombined to form representations of different specific events or of general events. A next object of study will be to specify the neural picture of memory construction, how neurons participate in different engrams and how retrieval transforms engrams. It seems that the same neural memory material is used for and participates in different mental episodes. It seems that this distributed participation produces variations both in the neural trace and in the mental representation. I shall study in particular the phenomena of engram co-allocation elsewhere.

Episodic memory does not guarantee content accuracy. This does not involve a direct realist view of memory, though, since memories are stored. But storage does not secure accuracy either, since it is a transformative process. The Lockean view of memory is thus unlikely: the preservative dimension of episodic memory is not the whole picture. Retrieval cannot present us with the ideas we originally experienced since there is a process of transformation of such ideas between the input and the output. Now one question is: what could a constructive memory theory of personal identity be? For now, it seems that a memory theory of personal identity endorsing a hybrid view of memory is better equipped to face Reid's objections. It can explain how personal identity can persist despite memory distortions and forgetting. I will propose a full defense of this claim in Chapter 5.

Section 6 – Concluding remarks

1/ The main problem of Locke's view is related to his definition of episodic memory, and in particular to (4) the sameness condition for remembering. Reid's objections are directed towards Locke's preservative view of memory.

2/ I have shown that there are two main strategies to solve the transitivity issue among Locke's scholars, but that both are missing a point made by Reid, against the sameness condition, and more generally accept a preservative view of memory.

3/ But there is a problem with the preservative view of memory. Memory cannot play the role of a faithful recorder in any theory, since it is not in fact doing exactly this job in real life. Relying on empirical evidence, I have defended two claims: witnessing an event does not secure the accuracy of memory recollection, and memory storage does not secure either the sameness or the accuracy condition. Thus, Locke cannot be saved from the second part of Reid's objection since it fairly attacks the preservative view of episodic memory.

4/ Locke is taken to give an objective account of personal identity, where the concept of personal identity can explain just punishment. A person would be justly punished if she is punished for something she in fact did. And just punishment is dependent on the truth of our judgements on personal identity. For episodic memory to play a constitutive role in personal identity it has to be accurate and to be a source of true beliefs. But episodic memory alone cannot always secure this. So, if Locke's account is an objective account of this sort, his account falls. But if we take the account in a more modest way, as saying something about how people can relate to themselves, and thus constitute themselves as extended selves, regardless of their accuracy, it might be different (Chapter 5).

5/ I do not deny the common claim that episodic memory generally allows me to know what happened to me in the past. But the traditional way of explaining it seems to rely on problematic presuppositions. Positing a difference in nature between false and genuine memories in virtue of the supposed absence of an engram in the case of false memories is for example subject to controversies. The question of the distinction between false and genuine memories and more generally the question of the constructive dimension of episodic memory will be the topic of Chapter 3.

6/ When and where the truth about personal past matters, episodic memory alone seems sometimes insufficient to access it. So that if we are seeking what persons truly did, episodic memory might not always be sufficient. It is likely that in some cases, memories need to be confronted with other sources to be disconfirmed or confirmed. This point, about how episodic memory, despite its constructive dimension, can be a source of self-knowledge, is in need of further inquiry (Chapter 5).

7/ If episodic memory sometimes looks like a weak epistemic source, it is unclear that the role of episodic memory for myself is only an epistemic role. A reflection on the diverse roles that episodic memory might play for persons will be the object of Chapter 4.

Second part
– The constructive dimension of
episodic memory

Chapter 3

– Are false memories memories? The conditions for remembering.

Introduction.

In the first part of the thesis, I have exposed a reading of Locke's theory of personal identity according to which personal identity depends on the epistemic access the subject has to herself. I have argued that memory is a capacity that both grounds self-recognition and self-extension to the past and constitutes personal identity between past and present selves. I have given arguments against the preservative view of memory and sketched a hybrid view arguing that episodic memory has two dimensions: it is both a preservative and a constructive capacity. I have suggested that such a view of episodic memory solves some issues of traditional memory theories of personal identity. But I have left a lot unexplained about episodic memory.

Philosophy of memory is a currently expanding field and I wish to devote the second part of this dissertation to the exploration of some of the issues being discussed in this field. The underlying belief is that a reflection on those is fruitful for the reflection on self-knowledge, self-extension and relationship to the past. Until now, for instance, I have left unsolved the question of what kind of mental

state can count as a memory. In the third chapter, I explore the conditions of remembering, and of successful remembering. Two main conditions are present in most philosophical accounts of remembering: the factivity condition and the causal condition. I argue that both of them carry presuppositions on the nature and functions of episodic memory: they presuppose that memory is primarily for the knowledge of the past.

In Chapter 3, I discuss those two conditions and show that their necessity for remembering is questionable.

In Chapter 4, I argue that episodic memory has also non-epistemic functions: memory is not only for the knowledge of the past. I mean that first, from a descriptive point of view, it does other things than preserving traces of past experiences, and second, from a normative point of view, we do not evaluate memories only as a function of how accurately they represent the past.

In the contemporary accounts of remembering, one major challenge is to distinguish among instances of memories and instances of other types of mental states; and among the former, a further challenge is to distinguish between instances of successful remembering and instances of unsuccessful remembering. Those two projects, although distinct, seem sometimes ambiguously treated. The case of false memory is a good illustration of this ambiguous treatment. If the very term of ‘false-memory’ suggests that those mental states are memories of a specific type (i.e. false), and that they are treated as such in memory science, some philosophers argue that false memories and other types of memory errors are not memories at all, because they do not fulfil the basic conditions for remembering. In this chapter, I discuss this claim.

The chapter goes as follows: In section 1, I describe a fictional case and propose four representations of the past that serve as a basis for the discussion. I then wonder about the reasons why some of them are considered as episodic memories and some of them are not. Building the argument on Bernecker’s distinction (Bernecker 2009) between truth (or accuracy with regard to the past event) and authenticity (or accuracy with regards to the past experience), I present the two main conditions for

remembering proposed in the literature: the factivity condition concerns the truth of the past event represented in a memory, while the causal condition concerns the authenticity of the past experience. In section 2, I focus on the factivity condition for remembering and argue that it is not really a condition for remembering but a condition for knowing the past on the basis of memory. I base the discussion on psychological research about induced false memories and the misinformation effect. In section 3, I focus on the causal condition for remembering. I base the discussion on the research about source monitoring errors. I argue that the causal condition is grounded on a misleading presupposition that a memory should and could have only one source: the past experience. I show that most accounts of remembering presuppose that episodic memory is primarily dedicated to the knowledge of the past, and, I argue, this conceals the variety of things that memory also does. I finally show that considering those two conditions as necessary conditions for remembering (instead of necessary conditions for successful remembering) leads to a paradoxical characterization of remembering: non-genuine memories are simply not memories at all, and unsuccessful remembering is impossible. But, as much as an account of representation should make room for misrepresentation (see for instance Nanay 2012) an account of remembering should make room for misremembering and other types of memory errors (Michaelian 2016c, 2016b; Robins 2016a).

Section 1 – A case.

Here, I propose a fictional case consistent with the research on false memory formation in which it is possible to identify at least one accurate representation of the past, and three distorted ones; two are inaccurate, and one is not. One is subject to a (mis)information effect from an untrustworthy source, another one is subject to an information effect from a trustworthy source; a third is due to a hallucination. The former two are both (partially) caused by post-event information. This case will

be a basis for the rest of the discussion, as it allows us to exemplify the range of various positions regarding the status of false and genuine memories.

Let's suppose that Anna has a memory of having killed Ms. F. She remembers giving her something to sleep because Ms. F was complaining about her sleeping issues. The next morning, Anna brings breakfast into Ms. F's room. She finds the body. Ms. F is dead, sudden cardiac death. On her nightstand, close to the bed, is an empty 2-pill package of hydroxyzine 50mg. When she sees them Anna starts to shiver. 50 mg is the maximum dose you can administer to an aged subject, Anna knows it. But the most common packages are composed of 25 mg pills, why is this one full of 50 mg? – what a horrible professional error. She remembers giving Ms. F sleeping pills last night, as she was on duty, and leaving the package on the nightstand (**Memory 1**). On the basis of this memory, Anna forms the belief that those are the pills she gave her last night. She hides the package in her pocket, leaves the room. She does her best to look quiet, head down. In the corridor, she sees a bin, and quickly throws the package away. She takes a glance around: no one. She looks back... damn, she sees a shadow -- uh, she sees someone -- in the door frame of Ms. J's bedroom. The old lady saw her, that is for sure. Weeks pass by, she feels more and more nervous and guilty about this. Three months after Ms. F's death, the case is reopened because of similar cases happening in the same institution. Investigation. Anna is consumed with fear and guilt. She cannot sleep anymore; she feels she carries an unbearable burden of guilt.. She decides to tell the truth. And, once in front of the detectives, she sincerely declares 'I remember I gave Ms. F Hydroxyzine to sleep. « What? », they say. Hydroxyzine, I remember looking for pills in the almost empty store, I saw hydroxyzine, we don't usually use them, but the store was kinda empty, and I took them, did not check the dosage. I gave her a glass of water; she took the pills. And never woke up. I killed her' (**Memory 2**). Next morning, I

found the empty package in the room and threw it away in the corridor, when I looked behind me, I saw Ms. J in her door frame, she must have seen me (**Memory 3**).”

Ms. J is interrogated; she did not see Anna in the corridor. The body is excavated, an autopsy happens. And no trace of hydroxyzine is found. But traces of a low dose of tryptophan, a nutritional supplement, used for sleep, but inoffensive at low doses. And another thing, they find a huge quantity of morphine in Ms. F’s body. Ms. F died from an overdose of morphine. At the end of the investigation, it appears that John, another nurse, was stealing morphine from the institution store and was administering patients morphine to euthanize them. It also appears that he put the empty package on the nightstand the night of Ms. F’s death in the place of the anti-allergic empty package to exculpate himself.

When Anna is given the new pieces of information, she realizes that she gave Ms. F tryptophane, she remembers going to the store, taking this, and giving it to Ms. F. When she tells the story after that, she *remembers* giving Ms. F anti-allergic pills for her to sleep--round, bigger and white (**Memory 4**). More importantly she does not remember being a killer anymore.

The first question I want to address is whether those four representations of the past can count as instances of remembering or not, and why. Bernecker 2009 has proposed the distinction between two meanings of accuracy in episodic memory. A memory can be accurate with regards to the past event, namely true; a memory can be accurate with regards to the past experience, namely authentic. According to him those two dimensions of accuracy are independent (see also Michaelian and Sant’Anna forthcoming).

There are two necessary conditions (among others) of remembering, present in most accounts, and that distinguish between genuine and false memories. One focuses on the truth of the retrieved past event and on the factivity constraint.

Factivity condition (FC): S remembers an event E if E. For a subject S to have a memory M of an event E, E must have happened in the past.

Another one focuses less on the past event as the input but more on the experiential basis of the memory and its causal relation with the memory⁵³. In this regard it is concerned with the authenticity of the past experience represented.

Causal condition (CC): For a subject S to have a memory M of a past experience Exp, Exp must have been operative in producing M.

On most accounts they are both necessary, but not individually sufficient (Bernecker 2009; Cheng and Werning 2016; Martin and Deutscher 1966). Here I will treat them separately and explore their individual consequences for an explanation of remembering.

Now let's give an analysis of the four memories proposed above on the basis of these two conditions for remembering.

M 1 (Anna remembers giving sleeping pills to Ms. F) seems to be able to fulfill both:

1/ **FC:** Anna in fact gave sleeping pills to Ms. F.

2/ **CC:** The experience of this last interaction caused her current representation of it.

M 2 (Anna remembers giving hydroxyzine to Ms. F) is a bit different, and could be described as an induced false memory:

⁵³ See Sant'Anna forthcoming for a discussion of Debus 2008's version of the experiential relational theory of remembering.

- 1/ **FC**: Anna in fact gave sleeping pills to Ms. F, but not these.
- 2/ **CC**: The current representation is not only caused by the past experience: Anna formed a memory on the basis of the integration of a new piece of information: the empty package of hydroxyzine on Ms. F nightstand.

M 3 (Anna remembers seeing Ms. J on her door frame) could be described as a memory of a hallucination:

- 1/ **FC**: Anna in fact had a hallucination. Ms. J was not on her door frame, but Anna mistook a shadow for Ms. J.
- 2/ **CC**: The current representation is caused by the past experience.

M 4 (Anna remembers giving Ms. F anti-allergic pills) could be described as a revised memory:

- 1/ **FC**: The current representation is accurate with regards to the past event.
- 2/ **CC**: Her current representation is caused by post-event information and maybe also by the recollection of the past experience.

	True (event) factivity condition: OK	False (event) factivity condition: NO
Authentic (experience) Causal condition: OK	M1	M3
Inauthentic (experience) Causal condition: NO	M4	M2

Table 1: Memory classification according to the factivity and the causal conditions.

On the factivity condition, Memory 2 and Memory 3 are not memories. On the causal condition, Memory 2 and Memory 4 are not. In section 2, I discuss the factivity condition, and in section 3, I discuss the causal condition.

Section 2 – The Factivity condition as a necessary condition for remembering

The factivity condition is considered as a necessary condition for remembering in most contemporary accounts (Bernecker 2009; Cheng and Werning 2016; Martin and Deutscher 1966). Here, I will take Cheng and Werning's 2016 as the main example since they are very explicit about it. On Cheng and Werning's 2016 account, factivity is a necessary condition for remembering. Memories 2 and 3 cannot be considered as memories in virtue of the missing factivity condition. They are false memories, and factivity being a necessary condition for remembering, they are not merely unsuccessful instances of remembering, they are simply not memories at all and do not count as memories. But what does that mean?

2.1. The analogy between fake guns, false money and false memory.

In a footnote, Cheng and Werning, 2016 draw an analogy between false memories, fake guns and false money. They argue that in those cases 'false' or 'fake' are privative. In their view, false memories are not memories, just as fake guns are not guns and false money is not money:

“The more plausible interpretation, we think, is that “false” in “false memory” is a privative adjective like “false” in “false money” or “fake” in “fake gun”. For privative adjectives, the inference is not valid: false money is not money, a fake gun is not a gun and, likewise, false memory is not a case of memory. Moreover, in psychological research and, even more so, in forensic situations the question of whether a memory report of the form “I remember that ...” truly is a case of memory or rather a case of confabulation or error often arises and is naturally regarded as a sensible question. This question would be pointless if memory were not generally regarded as factive. » (Cheng and Werning 2016, p. 6)

I first discuss the analogy, and then focus on the factivity constraint on remembering.

Following Dranseika 2020, I argue that their analogy is misleading. Nonetheless, the analogy that Cheng and Werning draw has the virtue of addressing an interesting question: what makes a mental state a memory? In virtue of what can we distinguish between money and false money, fake guns and guns, false memory and memory? The question of trait type individuation is a crucial question in the philosophy of biology, and there is an ongoing debate on how to individuate trait types (Burge 1989; Nanay 2010, 2012; Neander 1991).

I will try three criteria, proposed in the philosophy of biology to distinguish among them, and show that none of them can ground the specificity for all genuine money, memory and guns, and justify the analogy. As Nanay, 2010 has shown, there are in the philosophy of biology, at least three ways of individuating trait types, like hearts for instance. Biological traits can be individuated according to functional, etiological (Burge 1989; Millikan 1984; Neander 1991) or morphological (Nanay 2010, 2012) criteria. Nanay 2010, 2012, shows that none of them apply in all possible cases; he argues that the choice of the criterion depends on the explanatory project.

Similarly, I claim that the answer to the question ‘what makes a mental state a memory’ depends on the explanatory project one has in mind. What makes it possible to individuate money, guns and memory may also vary according to the explanatory context.

a. Functional criterion: First, it is not so clear that false in false money and fake in fake gun are privative adjectives. A straightforward way to define money, guns and memories is according to their function or role. Hearts pump blood (Nanay 2010), money pays, guns shoot, memories recollect the past. Their mock versions only *seem* to have the same role, but *do not*. You cannot buy things with false money, you cannot shoot with a fake gun, you cannot know the past with a false memory. At first glance. Under certain circumstances though, false money is money, and a fake gun is a gun, because they play the same role. I can rob a bank with a fake gun,

and I can buy bread with a false bank note, as long as they are really good imitations, and no one notices. Let's even imagine, that I do those actions without even knowing that they are neither genuine money nor genuine gun. It is likely that a lot of my memories are false, to some extent, though until I notice it, I will take them and use them exactly as I would use genuine memories, and they will play the same role, both in my mental life, in my social life and in communication. False memories can be constitutive for me of my representation of the past even if I mistakenly take them as accurate pictures of the past. They can be used as evidence for present beliefs about the past.

But there is something about fake news, fake guns, false money. They are not what they stand for. They are deceiving. They fulfill a role that they should not fulfill. On a functionalist definition, false memories could be memories, though they might not be instances of successful remembering. But what Cheng and Werning are interested in is not the question of what distinguishes successful and unsuccessful remembering, but what makes episodic memory a natural kind. Though, on a functional criterion only, there is not necessarily a distinction in kind between fake guns and guns, false money and money, false memories and memories. But then, how can we distinguish among them?

Let's go back to hearts (Nanay, 2010). All hearts pump blood is not enough to individuate hearts in virtue of their pumping blood. Is an electrical blood pump a heart? No, it can be an artificial heart, it could also be something very different. For an entity to be a heart according to a widely accepted theory of etiological function, it does not only need to pump blood, but it should be its 'normal evolved function' to pump blood 'in a body's circulatory system' (Burge 1989 cited by Nanay 2010, p. 415).

b. Etiological criterion: if not according their function, maybe we can distinguish them according to their etiology, namely according to their (evolutionary) history. In this regard, the analogy between false money and false memory might be more explanatory. The distinction between false money and money, is, at least according

to US law, in virtue of their production history, and more specifically, in virtue of who produces them. False money is produced by counterfeiters, money is produced by the governments, or the monetary authority.

But again, the same cannot be said of genuine and false memories; they can both be produced by the same person. In the case above, Anna produces each of the four memories, genuine and false. But one could say that their production history allows one in a sense to distinguish among them. They are all produced by Anna, but they are not all caused by her past experience. M2 and M4 are caused at least in part by post-event information. But Cheng and Werning, 2016 want to draw a distinction between false and genuine memories in terms of their truth, of their factivity, independently of whether they come from the past experience or not. Memories are supposed to be true. False memories are simply false. False money is simply false.

But for money it is a matter of convention which money is false and which money is genuine. It could be otherwise. Cheng and Werning, 2016 want to give an account of genuine memory as a natural (and not conventional) kind. But one could argue that the normal evolved function of memory is to represent the past accurately, thus, memory and false memory are of a different kind.

This would be an option, and the etiological criterion would seem to be better equipped than the functional one to specify what makes a mental state a memory. But I see three problems with this claim.

First, it presupposes that memory has been selected for truth preservation, though, one could argue that truth is a very recent concern compared with the capacity of remembering.

Second, the etiological criterion does not justify the analogy between false money and false memory. Even if one accepts that memory has been selected for truth preservation, the reasons to distinguish memory and false memory and money and false money are very different. The analogy in this regard might thus be a bit misleading, and the etiological criterion cannot ground similarly the distinction between money and false money and memory and false memory.

Third, even if one accepts that the etiological criterion is the best equipped to individuate memories, it is unsure that the view according to which episodic memory's normal evolved function is to preserve truth is the best way of understanding memory. On the contrary, as (De Brigard 2014) has shown, such a view on episodic memory's etiological function makes it difficult to understand how and why memory errors happen so often, and why in some cases they are beneficial.

The Deese-Roediger-McDermott (DRM) paradigm (Roediger and McDermott 1995), an experimental paradigm used for false memory research, is a good example of this. People are first presented with a list of words semantically related (e.g., *bed, rest, awake, tired, dream, wake, snooze, blanket, doze, slumber, snore, nap, peace, yawn, drowsy*). Then, they perform a recognition task: they are presented with words, and have to say whether those were present in the original list or not. Among them there are words that were in the original list, there is a *lure*, namely, a word which was not present in the list, but which is semantically related to the gist of the list (e.g., *sleep*), and some other words that are not semantically related (e.g. *candy, tree*). People typically do not make mistakes on the latter, though they recognize the lure as frequently as other words in the original list. If truth preservation is the etiological function of episodic memory, when someone recognizes the lure, one is not remembering.

Though such a view misses something in the description of this case. It is not the same to mistakenly recognize *sleep* and to mistakenly recognize *tree*. Recognizing *sleep* is best described as adaptive, recognizing *tree* is not. On Cheng and Werning's account, recognizing *sleep* misses the factivity condition, thus it does not count as an instance of remembering. However, when one recognizes *sleep* one remembers the gist of the list. This puts pressure on the claim that the etiological function of episodic memory is related to truth preservation. An alternative option proposed by De Brigard is to claim that the normal evolved function of episodic memory is to represent what could have happened.

c. Morphological criterion: a last option is that the analogy is justified on a morphological criterion, and that traits type individuation depends on morphological properties. The distinction between fake guns and guns, is, at least according to US law, in virtue of their mechanism. The US definition of a gun by law is as such:

“(3) The term “firearm” means (A) any weapon (including a starter gun) which will or is designed to or may readily be converted to expel a projectile by the action of an explosive; (B) the frame or receiver of any such weapon; (C) any firearm muffler or firearm silencer; or (D) any destructive device. Such term does not include an antique firearm.” (18 U.S. Code §921. Definitions).

Air soft fake guns expel projectiles by the action of compressed air; guns, by the action of an explosive. The distinction between them is thus legally drawn in virtue of their mechanism.

But the same cannot be easily said of genuine and false memories. There is an ongoing debate on whether genuine and false memories are produced by the same kind of mechanism, but it seems that they are hard to distinguish on either a phenomenal or a neural basis (for a discussion see Schacter and Loftus 2013, see also section 2.3.2).

It seems that Cheng and Werning’s analogy between fake guns, false money and false memory is at least a bit misleading. There does not seem to be a unified reason to distinguish between money and false money, fake guns and guns and memory and false memory. Moreover, the truth conditions of their claim, according to which fake guns are not guns, false money is not money, and false memory is not memory, depend on an explanatory context, and on one’s way to individuate trait types. However, Cheng and Werning’s project is to define memory as a natural kind,

namely, not in function of extrinsic and contextual properties, but in virtue of intrinsic properties. And one in particular: factivity.

2.2. The factivity condition

As the term ‘false’ in ‘false-memory’ suggests, false-memory seems at first to be a matter of accuracy regarding the past event. A false memory would be false because it is missing factivity.

This is Cheng and Werning’s suggestion. They write:

“Subjects frequently retrieve inaccurate information when asked to recall episodic memories. We regard these cases as improper episodic memory and aim for an analysis of episodic memory that presupposes its factivity”. (Cheng and Werning 2016:6).

Episodic memory, like knowledge, is often said to be factive. In epistemology, factivity generally refers to the presupposition of truth in propositions:

‘S knows that P’ is true only if P.

Some doubt that episodic memory has propositional content and prefer to talk about representational content in episodic memory (Michaelian, 2016c, p.69-70). I remain neutral on the nature of the memory content here and only argue that the idea that episodic memory is factive means that it presupposes the past event is true:

i. *Factivity condition with propositional content:* ‘S remembers that episode E’ is true only if E.

ii. *Factivity condition with representational content:* S remembers an episode E is true only if E happened in the past

So, memory factivity is intrinsically related to the crucial past event or past episode represented (or referred to) in a memory. Now, let’s explore what ‘if E’ means. They write:

“Our definition of episodes implies that an episode is an ordered list of events. As such, an episode is distinct from a set of events, which is an unordered list of events.” (Cheng and Werning 2016, p.7).

On Cheng and Werning’s view, the episode is the basis for the past experience but is distinct from it. So that, were this definition mixed with the factivity constraint on remembering, for a subject S to remember an episode E such as seeing a flash of light a short while ago, followed by a loud sound a few seconds later, an ordered list of events such as a flash of light followed by a loud sound few seconds later must be true.

2.3. Disjunctivism

2.3.1. Between genuine memories and memories of hallucinations

A first way the factivity condition has been criticized is to oppose arguments coming from the philosophy of perception. (Michaelian 2016c) for instance argues that the factivity condition is an entirely external condition and involves an unlikely disjunctivism.

“Consider a pair of simple cases. In the first, the subject satisfies whatever conditions he needs to satisfy in order to count as remembering, including factivity. In the second, he satisfies all conditions other than factivity, perhaps because he originally succumbed to an illusion and misperceived the relevant event. Given that we build factivity into our account of remembering, we are bound to say that he remembers in the first case but not in the second. Intuitively, this conclusion may sound right—in ordinary contexts, we do not tend to say that one can remember an event that did not in fact occur, any more than we say that one can perceive an object that is not there. But the conclusion comes at a high price. The only difference between the two cases, we

may suppose, is the accuracy of the subject's current representation, and this is purely a matter of a relation between the representation and past events. In other words, it is entirely external to the subject's cognitive processes and states. Psychologically speaking, everything is the same. » (Michaelian 2016c, p. 69).

Following Michaelian 2016c, let's imagine now that a subject S saw a flash of light and heard a quick sound but hallucinated a loud sound when she perceived it, just as Anna in section 1 threw the empty package into the corridor bin, saw a shadow in a door frame and hallucinated that Ms. J was watching. That allows us to understand a second problem with the factivity condition: as concerned with the truth of the past event, it is an objective and external condition. The factivity condition would have to determine objectively if a representation is true with regards to the past event to decide whether it is an instance of remembering. Though, for the rememberer, everything is the same. Anna remembers M3 just as she remembers M1. Her perception of the past might have caused her memory in each case. She remembers the past episode as she experienced it. The problem with M3 could be better described as a case where something went wrong during perception. But then the relation between the experience and the memory is a memory relation. The factivity condition presupposes a kind of disjunctivism – or a difference in kind -- between genuine memories and memory of hallucinations. This is unlikely, though, from an empirical point of view. Everything seems to be the same except the fact that the information retrieved is inaccurate in M3 (see also Perrin and Michaelian 2017).

2.3.2. Between genuine and false memories.

But this is not all. Even in the cases where nothing went wrong during perception, but the information remembered at retrieval is inaccurate, namely in the canonical cases of false memories such as M2, disjunctivism is unlikely. There is a huge literature on the study of the potential differences between memories and false memories, which pleads in favor of their phenomenal and neural indiscernibility (for a review, see Laney and Loftus 2013).

i. Phenomenal plan

It has been shown that remembering and misremembering can have the same extent of detail (Loftus 2005), and that they are phenomenally indistinguishable regarding their consequences for other thoughts and behaviors (most of the research on this uses food and taste, and shows for instance that false memories of food poisoning related to a specific food led to the diminution of their appeal and consequent consumption: see for example Bernstein and Loftus 2009; Geraerts et al. 2008; Scoboria et al. 2012).

They also seem phenomenally indistinguishable regarding their emotionality. Again, results suggest that emotion is not a good tracker of the difference. McNally et al. 2004, for instance, have shown that false memories of abduction by space aliens could not be distinguished from other traumatic memories the subjects had on the basis of their emotionality.

Furthermore, they appear to be indistinguishable regarding their endurance. Again, it has been shown repeatedly that false memories (and the distorted elements in them) can last as long as genuine memories (at least a year and a half; see for example Laney and Loftus 2008; Zhu et al. 2013).

The confidence people have in their memory has also turned out to be a bad tracker of difference since it does not seem to be correlated with truth or falsity (see for example Urgolites, Smith, and Squire 2015).

ii. Neural plan

The mechanistic decomposition enterprise has led to the same results. Studies have suggested a common network for remembering the past and imagining the past (Addis et al. 2012) and pointed out a common network for associative phenomena and genuine recognition (Gutchess and Schacter 2012). Neuroscience thus leads to the same results.

Although disjunctivism is a price that some agree to pay in philosophy of perception (see for instance, Soteriou 2000), so far, attempts to distinguish empirically between false and genuine memories have remained unsuccessful. Their difference in kind seems hard to account for.

2.4. Factive parts

A further argument against the factivity condition could be this: what if only one part of (a) is true and another one is not exactly true? Let's imagine that the subject S saw a flash of light, heard a sound but remembers that the sound was loud even though it was not. Anna in Memory 2 remembers parts of the episode properly, she is only wrong about one: the pills. But were she asked: 'how many pills did you give to Ms. F?' She could answer 'two' and would be right about it on the basis of her memory.

This might be a problem with the factivity condition: what if only some parts of the memory are not factive? It could be seen as an over-simplification to consider M2 as non factive since it has factive parts.

2.5. Misremembering

Furthermore, the distinction that the factivity condition involves between false and genuine memories hides the fact that memory accuracy can come in degrees. If one takes the factivity condition as a necessary condition for remembering, memories are always genuine, and false memories are not memories. It thus becomes impossible to account for unsuccessful remembering. But an account of remembering should make room for the distinction between successful and unsuccessful remembering (Michaelian 2016b; Robins 2016a). As Robins, 2016a has shown, memories can be inaccurate, but still result from retention processes. In this case, they could be better described as instances of misremembering or of confabulations.

2.6. What does it mean that a false memory is false?

Here I argue that there is a more general issue with the very term of false memory. ‘False’ in false memory, suggests that false memories are false with respect to the past event, that their problem is a problem of factivity. But what the research on false memories shows is that the issue in false memories is not the fact that they are false. In fact, in the typical cases of false memory formation proposed in the literature, the problem in the end is not that they are false; most of the time, the experiments do not help to decide whether they are true or false. They are said to be ‘false’ in virtue of the fact that they are caused by a misinformation effect. The term ‘misinformation effect’ in this regard carries with it the same kind of presupposition: integration of post-event information leads to false memories. In the following I show that those two terms are misleading. False memories produced in psychological research could be called inauthentic memories. And, as Michaelian 2013, 2016c has proposed, ‘misinformation effect’ could be replaced with the term ‘information effect’.

2.6.1. A problem with the term ‘false memory’

Memory 2 is typically described as a false memory. And this term seems to presuppose that false memories are false with regards to the past event. Such a presupposition is also carried by Cheng and Werning’s account: false memories are false because they are lacking factivity, namely, truth with regards to the past event. The term ‘false memory’, and especially in ‘false memory syndrome’ was born around 1992 (for a history of the term see Beckett 1996). As shown by DePrince et al. 2004, in a review of the literature on false memories since the 1990’s the term ‘false memory’, initially used to talk about (post-event) intrusions in memory, is widely used to refer to errors in details at retrieval. But still it has not and is not used to talk about memory of hallucinations such as in Memory 3. Memory 3 is not usually considered as a false memory in the psychological literature, but as a memory of a hallucination. I take this as a first set of evidence to say that false memories are not concerned with factivity, as Cheng and Werning suggested, but

concerned with causality. In this regard, ‘false’ in false memory is misleading since (despite what the term ‘false’ suggests) it refers more to inauthenticity than to falsity. What are considered as false memories are representations of the personal past that do not have the very past experience as their main or only source, and for which the past experience has not been operative in producing the correspondent mental state.

2.6.2. Typical cases of false memories in psychology as results of a misinformation effect.

The typical cases of false memories studied and produced by psychological research are induced by a misinformation effect. It has been defined as such:

“the impairment of memory of the past that arises after exposure to misleading information” (Loftus 2005, p. 361).

It gives rise to induced false memory. This phenomenon has been studied a lot in psychology, I will focus on two types of cases proposed in the literature:

i. Bugs Bunny case (Braun, Ellis, and Loftus 2002).

People are induced to remember a false detail in a memory of a real past experience. Before the experiment starts, while they are sitting in the waiting room, some of them can see an ad showing Bugs Bunny in Disneyland on the wall. Later, when asked about their memory of a day in Disneyland, some of them remember having spent time with Bugs Bunny on that day – this memory is impossible because Bugs Bunny does not belong to the Disney universe.

ii. Lost In the Mall (LIM) case (Loftus and Pickrell 1995).

People are induced to remember an entirely new past experience as having been lost in a shopping mall in their childhood. Similar types of settings have been used to induce a large range of false memories such as the construction of rich false memories of committing a crime (Shaw and Porter 2015).

In those two types of cases a subject can be induced to falsely remember an episodic memory or a part of it and imagine the false piece with increasing details (on imagination inflation: see Garry et al. 1996). LIM cases have been introduced as a really new piece of evidence: people could create entirely new false memories and confabulate or imagine lots of details about them (Loftus & Pickrell 1995) on the basis of suggestion.

But some of the induced false memories could very well be true. The issue of false memory is not about their truth or falsity but about their relation to the past experience. Let's imagine that in the Bugs Bunny case, the subjects of the study actually met Bugs Bunny when they went to Disneyland⁵⁴. Indeed, there were a lot of people at those times who loved to dress as Bugs Bunny while visiting amusement parks. And all of the subjects who remembered having spent time with Bugs Bunny actually were right. This might be highly improbable, but my point is that the experiment does not allow us to decide whether such memories are true or false. I don't claim that the memory is true, but that what the experiment really shows is that people can integrate pieces of information in their memories that come from different sources regardless of the truth of the information shared. It might be more accurate to talk about inauthentic memories than about false memories.

Such a shift in language might shed light in a dispute that started in the 1990's about false memory research, opposing false memory experts (especially Loftus), who were called as memory experts in trials dealing with recovered memories (often through psychotherapy) of sexual abuse in childhood on one side, and victims and psychologists on the other side. The debate has several aspects. One is on the existence and the possibility of true recovered memories of sexual abuse in childhood. As sometimes described (Crook 1995; Crook and Dean 1999), false memory experts tend to deny the possibility of true recovered memories. People

⁵⁴ This possibility is envisioned by Michaelian about the Lost in the Mall experiment (Michaelian 2016c, p.119). According to him, the subject remembers the experience regardless of whether she experienced it or not, as long as the memory is produced by a reliable episodic system of simulating the personal past. My argument is different in the sense that I do not use this case here to claim that a subject can be said to remember an event regardless of whether she experienced the event or not, but to claim that false memory research is not about the factivity condition.

have argued that the research on false memory tended not merely to show that some recovered memories could be false but more that all of them were false (Crook 1995; Crook and Dean 1999). Bernstein and Loftus write for instance: ‘In essence, all memory is false to some degree’ (Bernstein and Loftus 2009, p. 373). Another aspect of the dispute is about the consequent devaluation of victim’s--and especially women’s--voices in this kind of trial. The reproach is that this underlying claim about the impossibility of true recovered memories of sexual abuse in childhood tends to distrust people that have already less credibility especially in Court (Sexton, Chrissy, 2018), to undermine speakers that are already seen as less competent (see for example Sellnow and Treinen 2004), and to undermine memories that are already repressed. If, as I have argued, false memories deal with the causal grammar of silence condition on remembering, the debate might be different. False memory research cannot decide whether memories are true or false but can raise doubts on their causal history. This leads to an exploration of the conditions of integration of new pieces of information. In the rest of the section I follow this lead.

Another interesting lead that the dispute on false memories opens concerns the social and cultural conditions for remembering. In a different intellectual context, the anthropologist Dussy 2013 investigated incest with the following interrogation: why are memories of sexual abuse in childhood massively repressed? She makes the hypothesis that incest in particular is protected by the system of silence. A “grammar of silence and domination”, in her terms, surrounds it and makes it impossible for the victim to talk about it or remember it. In a more positive context, I remember X saying, after a great afternoon ‘we will remember this’, prompting both of us to remember. Memories in each of those cases might have been repressed or prompted by a cultural and social environment. I will leave this lead aside for now.

2.7. Disentangle post-event information integration and falsity.

The integration of post-event information after exposure to a misinformation effect has been described as a robust phenomenon, observable also in ‘real life settings’

(Seamon et al. 2006). It tends to produce a picture of the episodic memory system as a system highly vulnerable to exposure to exogenous and misleading information (Conway and Loveday 2015).

The research on false memories and misinformation effect tends to claim not only that:

- (a) If the cause of M is not Exp, then M is inauthentic.

But also, that

- (b) If the cause of M is not Exp, then M is false.

Thus, it tends to present the factivity condition and the causal condition as interdependent. Either they are fulfilled together, or they are unfulfilled together. I contend that this is misleading and discuss (b). Following Michaelian 2013, I argue that the vocabulary of misinformation effect is misleading, because it presupposes that post-event information integration in an episodic memory results in content falsity with regards to the past event. This is not necessarily so, though.

2.7.1. Representations of a past event caused by the past experience can be false.

The fulfilment of the causal condition does not necessarily entail the fulfilment of the factivity condition.

First, an inaccurate (false) representation of a past target event can arise out of a distortion in memory reconstruction of an accurate representation. But it can also arise out of an undistorted recollection of an inaccurate representation. Indeed, memory construction does not happen only at retrieval but also at encoding (Campbell 2014; McCarroll 2018). Memory 3 is a good example of that: Anna remembers seeing Ms. J in her frame door. This inaccurate piece is not added at retrieval, but she constructed this representation at encoding.

If episodic memory is constructive also at encoding, there is room for cases where post-event information integration allows one to form a more accurate (true)

representation of the past. Thus, a distortion of an inaccurate representation could also in principle lead to an enhanced accuracy of the representation of a target event.

2.7.2. Integration of post-event information into a memory can lead to an accurate (true) representation of the past event.

The unfulfillment of the causal condition does not necessarily entail the unfulfillment of the factivity condition.

i. Information effect.

In his 2013 paper Michaelian envisions cases where post-event information is true. While he acknowledges that psychology has focused on the integration of inaccurate post-event information, he focuses on the integration of accurate post-event information.

Memory 4 is a good example of post-event information leading to an accurate (with regards to the past event) memory. Anna revises Memory 2, on the basis of post-event information, and the resulting representation of the past is accurate with regards to the past event. In other words, it is true. The conditions under which a memory distortion produced by post-event information integration leads to an accurate memory or not depend on the quality of the information integrated. But the integration of post-event information into a memory does not necessarily lead to its inaccuracy. Following Michaelian 2013, 2016, I think we should replace the term ‘misinformation effect’ with the term ‘information effect’ to disentangle post-event information integration and falsity.

ii. Epistemic contexts of the integration of the new piece.

Michaelian 2013’s main claim is that the memory beliefs resulting from an information effect where the integrated new piece is true, can qualify as knowledge as they are not in general luckily true but produced by a reliable memory system. According to him, the constructive view of memory is compatible with the idea that the episodic memory system is directed towards the knowledge of the past. Here I

argue that even though the research on false memories and misinformation effect produces a picture of an unreliable memory system, it could also be seen as evidence for the claim that the integration of post-event information depends on individuals' specific epistemic contexts. This could give further arguments to Michaelian's reliabilist theory of episodic memory and especially to the claim that episodic memory construction is dedicated to the knowledge of the past in a collective epistemic enterprise. I will discuss this view in section 4.

In the first type of case, the false memory is induced by an advertising context in a waiting room. Here the subjects integrate new pieces of information in a trustworthy perceptual context. This environment is not supposed to be misleading. The advertisements seen in this context, even though they might be considered as manipulative by the perceiver, like all advertisements, are not supposed to be aiming at deceiving her.

In the LIM experiment context, one episode among four concerning one's childhood is false and inauthentic. The inaccurate episode nonetheless relies on detailed information from relatives. The three others are true events that have occurred between the subject's 6th to 8th year, reported by relatives. Here is an example of a false piece given to one subject in Loftus & Pickrell, 1995:

'You, your mom, Tien and Tuan, all went to the Bremerton K-Mart. You must have been five years old at the time. Your Mom gave each of you some money to get a blueberry Icee. You ran ahead to get into the line first, and somehow you lost your way in the store. Tien found you crying to an elderly Chinese woman. You three then went together to get an Icee'. (Loftus and Pickrell 1995, p. 721).

When the subject remembers this false piece, they say:

« I vaguely, vague, I mean this is very vague, remember the lady helping me and Tim and my mom doing something else, but I don't remember crying. I mean, I can remember a hundred times crying... I just remember bits and pieces of it. I remember being

with the lady, I remember going shopping. I don't think I. I don't remember the sunglasses part. » (Loftus and Pickrell 1995, p. 723).

After the debriefing, when the subject is informed what this study was about, and asked to guess which piece is the false piece, here is an example of how they react:

“Well, it can't be Slasher, cause I know that he ran up in the chimney and I know that that car got smashed and I know that we got robbed so it had to be that mall one.” (Loftus and Pickrell 1995, p. 723).

According to Loftus & Pickrell sometimes people don't identify the false piece when they are asked to identify it after the debriefing. But when they succeed, it's often only by elimination. On Loftus and Pickrell, this is a sign of the fact that episodic memory is highly vulnerable to suggestion and induction. But again, here, there is a context of the integration of the false piece. Subjects are presented with narratives of their childhood; they ignore the topic of the experiment; they are not suspicious. A trustworthy source of information, who is working in a lab, who apparently possesses accurate and detailed knowledge of their past (thus who is likely to appear as reliable) is burying a false episode among true pieces.

The typical cases of false memories in psychology are cases where subjects are induced by various means to remember false pieces of information, and where evil demons voluntarily induce subjects to integrate information with the intention to deceive them.

But we could also see the picture as such: one thing that emerges from false memory research is that subjects have trust in the advertising context of a hospital waiting room, in the psychologists, in their relatives, in a group of peers (Bruck, Ceci, and Hembrooke 2002). It is unsure that those studies do show that in any context one can remember just anything. They could only show that subjects can integrate pieces of information in their own memories depending on an epistemic social

context: remembering is a collaborative enterprise (Harris et al. 2019; Meade et al. 2017).

In some cases, memory distortion leads to enhanced knowledge of the past. Some constructivists argue that the constructive dimension of episodic memory is functional in this regard (De Brigard 2014; Michaelian 2016c). Some others argue that memory distortions should be explained in terms of background beliefs and hypothesis confirmation (Mahr and Csibra 2018). I will discuss these views in the next chapter. For now, I will remain neutral on this issue. For the current argument, it is enough to notice that false memory research is not about the factivity condition and that integration of post-event information into a memory does not necessarily lead to its inaccuracy.

2.8. Concluding remarks on factivity

1/ Psychological research on false memories does not allow us to decide whether false memories are true or false with regards to the past event. The problem with the misinformation effect and false memories is more related to the causal condition.

2/ The causal condition and the factivity condition are presented as being interdependent. But they are not necessarily fulfilled together. The fulfillment of the causal condition does not entail the fulfillment of the factivity condition and the unfulfillment of the causal condition does not necessarily entail the unfulfillment of the factivity condition.

3/ In fact, the main reproach one could address to Cheng and Werning's account, and more generally to the factivity condition for remembering, is that they do not answer the question: 'what are the conditions for remembering?'. They answer the question: 'what are the conditions for knowing the past on the basis of remembering?'. The factivity condition seems to answer a partial question about remembering which is: what are the conditions for knowing the past on the basis of memory? Building their account on the rejection of false memories, on the basis of

the factivity condition, Cheng and Werning, 2016 actually convey an interpretation of the term false memory which is not in line with its use in psychology. Most of the time, psychological research on false memories does not allow us to decide whether false memories are true or false with regards to the past event. The problem with the misinformation effect and false memories is more related to the causal condition.

4/ A related issue of the accounts that consider the factivity condition as necessary for remembering is that they might be unable to make room for misremembering. On their accounts misremembering simply is not remembering. So that they cannot explain memory errors. Memory errors on their view are merely errors.

Section 3 – The causal condition as a necessary condition for remembering

Contrary to the factivity condition, the causal condition for remembering is more focused on the past experience as the memory input. On the causal condition, M2 and M4 cannot count as instances of remembering in virtue of the fact that they are not caused in an appropriate way by the past experience. This is regardless of whether they are true or false. The matter with M2 and M4 according to a causal theory of memory is that they are inaccurate with regards to the past experience. They might be true, but they would still be produced by an improper source, and thus, depending on the accuracy of the resulting representation, could count as instances of knowing but not as instances of remembering. In this section, I propose to contribute to the debate on the causal theory of memory initiated by Michaelian 2011, 2016c and De Brigard 2014. I explore what it could mean for a memory to be caused directly by the past experience and contend that both encoding, and retrieval correspond to the gathering of information from various sources. In this sense, the causal theory of memory is either too loose or too strict. According to

Michaelian, 2016c, once we understand episodic memory as a constructive process, we are pushed to abandon the causal theory and to endorse a simulation theory of memory. I will discuss this claim in Chapter 4.

3.1. The causal theory of memory

The factivity condition on remembering is too strict, because in particular, it puts truth as a necessary condition for remembering, even though one could be wrong about what happened but still be right about how she experienced the event. The factivity condition thus puts the stress on the truth of the past event or episode regardless of the accuracy of the memory with regards to the past experience.

The causal condition of remembering on the contrary puts the stress on the past experience. There is a current debate in philosophy of memory on the causal condition of remembering, opposing causal e.g. (Bernecker 2009; Martin and Deutscher 1966) and post-causal (Michaelian, 2016c; De Brigard 2014) theories of memory (for an exposition of the dispute, see Michaelian and Robins 2018; Perrin and Michaelian 2017; Robins 2016b). The main argument against the causal theory of memory is that episodic memory can produce both new beliefs and new content, and the causal theory has difficulties accommodating this (Michaelian, 2016a). The causal theory of memory has been introduced by Martin & Deutscher, 1966, even though it is presupposed in earlier theories such as Locke's, for instance, and is widely accepted in philosophy of memory (apart from few exceptions such as Michaelian, 2016a and De Brigard 2014). The idea is this:

Causal condition: For a subject S to remember a past experience Exp, the past experience must be operative in producing (Martin & Deutscher, 1966) –or must have caused with an appropriate causal link-- the memory M.

This condition is considered as a necessary condition, though not as a sufficient one. As in Cheng and Werning, 2016 (who posit both the factivity condition and the causal condition as necessary for remembering), in Martin & Deutscher, there are

further conditions for remembering, and factivity is one of them. It is not enough that a memory is caused by a past experience, it also has to be an accurate representation of the past event:

*“Within certain limits of accuracy, he represents that past thing.”,
Martin & Deutscher, 1966, p. 166.*

The accuracy they refer to here is accuracy with regards to the past event. As argued in section 2, this further condition is thus another formulation for the factivity constraint on remembering. I have already discussed the factivity condition. Let's focus here specifically on the causal condition and its consequences.

On the basis of the causal condition only, Anna's...

1. Memory 1 can be an instance of remembering. It can be described as caused by the past experience it aims at representing.
2. Memory 2, though, cannot. It is both false with regards to the past event and inauthentic with regards to the past experience. It does not count as an instance of remembering: it is not caused by the past experience but by misleading post-event information.
3. Memory 3 could be an instance of remembering. Even though it is false with regards to the past event, it might be authentic with regards to the past experience: even though she hallucinated, Anna had the experience of seeing Ms. J in her doorframe. Her memory could be caused by this past experience.
4. Memory 4 is true, or accurate with respect to the past event. But is it caused by the past experience in an appropriate way? To answer this question, we need to understand what appropriate means in a causal theory.

And what the appropriateness of the causal link exactly means is not very clear. Here, I take that by 'appropriate' Martin & Deutscher mean 'non-deviant':

Appropriate causal link: S's memory M of the past experience Exp should be caused by a non-deviant causal chain.

Memory 4 is quite similar to a case they draw in their 1966 paper and reject as a case of remembering. But in a sense their case makes it clearer that they mean ‘direct’. They imagine this:

“Kent was involved in [a first] accident and as a result told Gray about it. Because he has been told about it, Gray related the story back to Kent, after Kent had lost his memory in [a] second accident. Thus, Kent’s original observation of the accident was a factor in bringing about his final account of it. His observation of the accident, however, is not operative in producing (through a successive set of states) his account of the accident in the circumstance of his being prompted. For that reason we do not say the he remembers the accident.” Martin & Deutscher, 1966, p 186.

Anna’s memory (in Memory 4) could be seen as caused in a way by her past experience: one could argue that the results of the investigation trigger her memory and make her retrieve it, and I will come back to this hypothesis in section 4. But according to Martin and Deutscher, if Anna’s representation of the past does not contain any more details than what has been prompted, she cannot be said to be remembering. Indeed, her past experience, given that she has been totally prompted (let’s assume that at the end of the investigation she is given all the details she has retrieved in Memory 2, plus the detail of the exact pills she gave to Ms. F, including their name, shape, color and composition), cannot be said to be operative in producing her representation of the past. In other words, as the past experience is not necessary for Anna to form the representation of the past that she now has, we can say that she knows but we have no evidence for saying that she remembers.

In Kent’s case, it is slightly different, and easier to argue for since he is described as suffering from an episodic amnesia. Martin and Deutscher argue that he is not remembering. His representation of his past experience even though accurate with regards to the past event, is not directly caused by the past experience but indirectly

through Gray's testimony. It is caused by post-event information on this past experience.

On the causal theory, both Anna in Memory 4 and Kent are not remembering. They believe they are, but they are committing a source monitoring error (Johnson et al. 1993). Namely, they form a mistaken metacognitive judgement about the source of their representation of the past and think it comes from their past experience. But the source of their representation is post-event information.

The focus of this section is thus on Memory 2 and Memory 4. Those two representations do not count as instances of remembering on many accounts because they do not fulfil the causal condition of remembering. The causal condition of remembering is thus used to distinguish between a lucky authentic representation of a personal past experience and a genuinely authentic memory. It leads to excluding two types of representations of personal past episodes from the category of episodic memory:

- i. Those that are inauthentic and false and produced by a misinformation effect (as it is called in the literature), such as Memory 2;
- ii. Those that are true but still result from a source monitoring error, such as Memory 4, and thus are said not to be authentic either.

In this section I go on to discuss the causal condition on remembering, and the terminology surrounding episodic memory errors. 'Source monitoring error' (Johnson, Hashtroudi, and Lindsay 1993) presupposes that a memory should have only one source (the past experience) and that the source could be identified. I contend that this is misleading. I eventually argue that the vocabulary surrounding episodic memory misleadingly presupposes that episodic memory is only busy with rejecting endogenous information, preserving the authenticity of the past experience, or searching for it. I contend that the binarism between false and genuine memory is not helpful to explain episodic memory. On the contrary it posits too strong constraints on remembering, such that few representations of a personal past experience could count as successful remembering.

3.2. The boundaries of the past experience.

Here, I discuss (a):

If the cause of M is not Exp, then M is inauthentic

3.2.1. Remembering more – inattentive encoding

What does it mean for a memory to be authentic with regards to the past experience? People have argued that one was encoding more information than what she was consciously aware of during the past experience. According to Boyle 2019, episodic memory is epistemically generative, namely, one can know more about a past event on the basis of an episodic memory than what one explicitly knew at the time of the event.

She gives two types of cases to support this claim: the case of the taxi driver she borrows from Lackey 2005, and the case of the bite.

In the latter, the rememberer learns from the past because she received post-event information on the meaning of a word:

“As a child of about six, I announced to a group of adults ‘I’ve been bitten by a midget!’, believing this to be the term for a small biting insect. I could not understand why they laughed. Years later, I remembered this episode and knowing, by then, that ‘midget’ is a somewhat offensive term for a person of unusually short stature, I laughed too. I knew at that point, on the basis of memory, that I had said something ridiculous” Boyle 2019, p. 2.

This leads her to revise her memory. To come back to the terminology I used in this section, I would say that this memory is reconstructed and true as it is accurate with regards to the past event. The rememberer knows more about the past event thanks to post-event information.

The former case, the taxi driver case, is slightly different:

“Yesterday morning was like most others for Clifford: he spent [it?] drinking coffee, listening to the radio and driving his car during his hour-and-a-half commute to work. As was typical for these commutes, Clifford’s attention was divided between the other cars on the road, the surrounding environment, the discussion and music on the radio, and his thoughts about the day’s work. Because of this perceptual and cognitive overload, Clifford found himself, as he often did on these drives, taking in more pieces of information than he actually processed at that time.

Indeed, this was made apparent earlier this morning, when Clifford bumped into his friend, Phoebe, at the bakery and started talking about his commute. During the conversation, Phoebe asked him whether construction had begun on 155. (...) Upon being asked this question by Phoebe, Clifford paused, called to mind passing 155 on his drive to work yesterday, and correctly remembered seeing construction work being done on this freeway.” Lackey, 2005, p. 650.

In this case, Clifford learns about his past experience on the basis of remembering. This leads to another point: not only can remembering generate knowledge about the past, but also one can remember more of an experience than one was previously aware of while having this experience. But then, here is a puzzle for the causalist. On the causal theory, the past experience has to cause the memory in a non-deviant way. Does Clifford remember? Either the causalist answers that he does, in which case she would face a first issue. Indeed, she would have to endorse the controversial claim that experience can be unconscious (few people are willing to endorse this claim (though, see for instance Rosenthal 2009) but it is highly controversial). Or, the causalist answers that Clifford does not remember. And then, she would face another issue. Clifford’s memory is not exactly inauthentic, as it is not exactly unrelated to his past experience.

Either the causal theory accepts that the past experience is larger than the conscious experience, and then it might be too loose. Or it denies that the past experience has non-conscious parts. And in this case, it cannot explain how Clifford is remembering.

3.2.2. Remembering more – filling gaps

The claim (a) relies on a presupposition that for memory to target the past and to be accurate, it has to be caused by the past experience. Only in that case is a representation of the past a memory. The causal condition is supposed to secure memory accuracy. If one remembers more than what one previously experienced, one's memory will be inaccurate. But this is unsure, and such a view leads to a weird description of remembering. We often remember more than what was present in the past experience, and episodic memory in this sense seems to be regularly filling gaps. And filling gaps does not systematically lead to inaccuracy. Thus, such a view would have two difficult consequences. If we examine again the DRM paradigm for instance (mentioned in 2.1.b), it appears that filling gaps is both consistent and not dysfunctional. So, first, if filling gaps is not what memory is supposed to do, why is it so consistent? Second, if this is dysfunctional, how do we explain why filling gaps often leads to accurate representations of the past? Plus, how can we explain why people with episodic amnesia make fewer recognition errors in the DRM paradigm than healthy subjects as shown by Schacter et al. 1996 (see also De Brigard, 2014)?

3.3. The past experience as the source -- Source monitoring error

Memory 4 could be seen as very similar to Clifford's case. But what prevents causalists from saying so is that it is unclear whether Memory 4 is caused through a non-deviant causal chain by Anna's past experience or caused by post-event information. She might be making a source monitoring error: believing she remembers when she just knows by external sources. But what I want to argue for

here is that the terminology of source monitoring error presupposes that an episodic memory should have only one source, the past experience, and that this source could be clearly identified. Here, I discuss the relation between perception and memory through the phenomenon of boundary extension, where people believe they remember a scene they just saw but in fact remember more than what they just saw. As they extend the boundaries of the scene by imagining, and do not identify this added part as coming from their imagination, they make a source monitoring error. I argue that perception normally relies on object completion. I then explore the relations between semantic and episodic memory through the experiments in face recognition. Episodic memory, as the recollection of a past event perception, cannot do better than perception of the past event. If representing a scene through the perception of it relies on some other sources than the very scene perceived, the corresponding episodic memory cannot either be sourced only in the past experience. I show that episodic memory depends at encoding and at retrieval on completion processes. In this sense the source of an episodic memory cannot only be the past experience it aims at representing.

3.3.1. The phenomenon of boundary extension and the gathering of information from various sources at encoding.

In the case of boundary extension, people remember more than what they just saw. They are presented with an image for a short time, and as soon as 42 ms later (Intraub and Dickinson 2008), they are presented with two images: one which is the same as the one they just saw, one which has extended boundaries. When people recognize the extended one, they are demonstrating boundary extension (Intraub and Richardson 1989).

It has been shown that the boundary extension phenomenon is attenuated in people with episodic amnesia (Jajdelska et al. 2019; Mullally, Intraub, and Maguire 2012). People with episodic amnesia correlated with hippocampal lesions remember scene boundaries more accurately than controls. Amnesiac patients seem to have general difficulties with spatial coherence and a correlated attenuation of the expression of boundary extension (Chadwick, Mullally, and Maguire 2013).

Though the attenuation of the boundary extension phenomenon in amnesic patients might not be sufficient evidence to claim that episodic memory is involved in boundary extension. Another explanation for this fact would be that the hippocampus is not only involved in episodic memory but also involved in spatial imagination, and in some aspects of perception. This has been argued for example by Maguire, Intraub, and Mullally 2016. A study by Mullally, Vargha-Khadem, and Maguire 2014 seems to support this: they showed that in developmental amnesia, the boundary extension phenomenon was observed. It suggests that it is partly independent from episodic memory.

Nonetheless the boundary extension phenomenon is a very robust effect observed in infants as young as 3 to 4 months old (Quinn and Intraub 2007), in children (Kreindel and Intraub 2017) as young as 6 years old (Candel et al. 2004), and in adults as old as 84 (Seamon et al. 2002). Such an error of commission can happen as quickly as following a 42 ms break in the sensory input (Intraub and Dickinson 2008). But how can it happen so robustly and quickly? A hypothesis is that it is because perceiving involves the gathering of perceptual, episodic, spatial and semantic information. If boundary extension has been shown as partly independent from episodic memory, it might not be independent from semantic information. Some have argued that we form representations of the world based on semantic information or learned categories (Intraub 2012). Mamus and Boduroglu 2018 have hypothesized for instance that the perceptual context was playing a major role in the boundary extension phenomenon. In an experiment, they used abstract shapes instead of identifiable and meaningful objects. For abstract shapes, no boundary extension effect is observed. When the shapes are not semantically identified, the phenomenon of boundary extension does not arise. According to them, this pleads in favor of the fact that boundary extension is correlated with a construction of a representation on the basis of the scene observed and on the basis of related information. It is unclear though whether it shows that the phenomenon of boundary extension arises because of the involvement of semantic information during perception or whether it only involves object completion (McDunn, Siddiqui, and Brown 2014; Nanay 2018b). Anyway, what it certainly shows, is that perception involves seeing more than what is present to my sight in the world, and this leads to endorsing the information gathering claim:

*Information*⁵⁵ *gathering claim*: A's representation of a scene is not only sourced in what is available in the experience of the scene but depends on the gathering of information from various sources.

If a representation in perception is not only sourced in the very experience it aims at representing but depends on the gathering of information from diverse sources, if encoding is dependent on perception, and if retrieval is dependent on encoding, it is unlikely that the source of a memory at retrieval is only the past experience. Here is a new puzzle for the causalist:

If (a) the information gathering claim is true (and it seems to be supported by empirical evidence), if (b) encoding is dependent on perception (uncontroversial), and if (c) retrieval is dependent on encoding (accepted by the causal theory of memory), it follows that it is unlikely that the only source of a memory at retrieval is the past experience.

The representation of a scene seems to extend beyond what is available in experience. So, the retrieval of such a representation is unlikely to be sourced only in the past experience. On the causal condition, genuine memories are caused through a non-deviant causal chain by the past experience they aim at representing. But the research on boundary extension casts doubts on such a view of remembering. Encoding depends not only on what is available in experience but also on object completion. If encoding is not directly caused only by what is available in experience, how could retrieval be?

I argue further that retrieval also involves the gathering of information from various source.

⁵⁵ I here use a folk concept of "information", and refer to information in a weak sense here, and not to refer to semantic information.

3.3.2. Interdependence of episodic and semantic memory at retrieval and conclusion on the causal condition.

As argued in Chapter 2, episodic memory has a constructive dimension and preservationism is unlikely. There was a debate on the interdependence of episodic and semantic memories in the 1980's. While Tulving 1984 suggested that episodic memory is contingent on semantic memory, Squire 1992 for instance argued that as they were selectively impaired, they were totally independent. But more recently, growing evidence has shown that episodic memory construction at least partly depends on semantic information. It has been shown that semantic and episodic memory tasks were involving common brain regions (Barba et al. 1998). Further, it has been shown that episodic retrieval also was dependent on semantic information. For instance, in La Corte et al. 2012 the relation between semantic and episodic memory with face recognition for famous and unknown faces was investigated. They compared semantic encoding for faces to their recognition in episodic memory. The results show that the level of semantic knowledge of faces affected their recognition in the episodic memory task. People were more likely to recognize famous faces than unknown faces, suggesting that episodic recognition is dependent on semantic information.

3.4. Concluding remarks on the causal condition.

1/ Those studies altogether seem to show that episodic memory is not an independent process but interacts with other types of information. So that it makes little sense to say that a retrieved episodic memory comes only from the very past experience it aims at representing. On the contrary, it seems that episodic memory recollections also take their source in semantic memory. As semantic memory is formed from learning and generalization of specific past experiences (Tulving 1985), saying that episodic memory relies on semantic information involves that a specific episodic memory cannot only depend on what it aims at representing but also on other memories and other past experiences. In this regard the causal condition on remembering is misleading. Since retrieval depends on the gathering of information coming from diverse sources, saying that a memory should be caused by the very past experience it aims at representing is an oversimplification.

It is unclear whether any representation of the past can entertain such a causal relation to the past experience.

2/ Either the causal condition is a loose condition according to which remembering Exp has to be caused partly by Exp, but then it is not very explanatory. Indeed, on the basis of such a loose causal condition, each of M1, M2, M3 and M4 can count as memories: they can all be described as partly caused by the past experience they aim at representing. M2 could be caused partly by the experience of putting the pills on the nightstand, M4 could be caused partly by the experience of giving pills to Ms. F. Such a loose condition loses its explanatory power. Or the causal condition is a strict condition, according to which remembering Exp has to be caused (through a non-deviant causal chain) by Exp, and then it does not fit with memory science. Neither a representation of Exp nor a memory of Exp can be caused (through a non-deviant causal chain) by Exp. They are the result of the gathering of information from diverse sources. On a strict causal condition none of M1, M2, M3 and M4 can count as memories: they can all be described as caused by the gathering of diverse sources of information. Such a strict condition would also deprive the causal theory of explanatory power, since it cannot account for a distinction between misremembering and imagining the past.

3/ The causal condition thus has a descriptive problem. Remembering is more likely to be the construction of a representation of the past on the basis of the gathering of information from various sources. An alternative to the causal theory has been proposed by Michaelian, 2016c: Remembering is the simulation of one's past.

4/ But the causal theory had other ambitions beyond only descriptive ones. It was supposed to explain how memory could be a source of accurate representations of the past. This also is the ambition of simulationists. According to the causalists, the episodic memory system is a system that represents the past on the basis of a memory trace initiated in the past experience. According to the simulationists, episodic memory generally is a system that reliably simulates the past.

Section 4 – Concluding remarks

1. The factivity condition presupposes that episodic memory's function is related to truth preservation. This is unlikely.
2. The causal condition posits that for a memory to be an accurate picture of a past experience, it has to be caused through a non-deviant causal chain by such an experience. This is unlikely too. Memories have various sources. And representations of past experiences can be reconstructed from those various sources but still be accurate.
3. A theory of remembering should make room for errors. If the factivity and the causal conditions are necessary conditions for remembering, then, one can only successfully remember. But, as much as one can misrepresent but still represent, one can misremember.

Now, I intend to show that beyond their individual difficulties, the factivity condition and the causal condition, by focusing on the truth of the past event or the authenticity of the past experience in episodic memory, presuppose that memory is primarily busy with rejecting exogenous information. Those conditions imply a kind of preservationism, which is unlikely. Moreover, this hides the variety of things that the episodic memory system is busy with. And in a sense, this reproach is more generally addressed to any epistemic theory of memory, including simulationism. In particular, how can they explain the common phenomenon of forgetting? On these accounts forgetting has to be dysfunctional. I contend that this might be misleading too.

Chapter 4

– Non-epistemic memory evaluations and non-epistemic memory functions.

Introduction

People have recently argued that while focusing on memory errors philosophers of memory fail to see what memory is for: knowledge of the past. According to (Mahr and Csibra 2018) for instance, we use memory in communication to warrant the fact that we are a trustworthy source of information about the past. They argue that the descriptive project that puts the stress on memory errors is separate from a normative project concerned with memory evaluations. One central question about memory is concerned with memory evaluation and, they claim, memories are evaluated in function of their accuracy and their epistemic value about the past. For this reason, continuism (the position that holds that episodic memory and episodic future thought are of the same kind; the terminology is from Perrin 2016; for a discussion of continuism and discontinuism see Perrin and Michaelian 2017; for a defense of continuism see Michaelian 2016a, for a defense of discontinuism see Robins 2020) is not justified. The empirical project and the normative project are two separate projects, focusing on two different aspects of episodic memory. They argue that the empirical findings on the constructive dimension of episodic memory do not change a trivial fact: memory is for the knowledge of the past. It is how we evaluate memories and it is related to episodic memory's main function. In this chapter, I discuss these claims. I start with a presentation of the justified concerns

about the reduction of memory research to the descriptive project and discuss the normative one (section 1). I contend that the epistemic norms of memory evaluation are less unilateral than assumed by the normative view. In particular, I propose that epistemic norms depend on specific contexts of evaluation. I discuss Craver 2020 analogy between episodic memory and the game of chess and discuss the case of confabulation. Building the argument on the recent research on the epistemic benefits of memory distortions in clinical confabulation, I propose that when the usual epistemic norms cannot be fulfilled some other norms may apply (section 2). Then, I go on towards a more radical view and propose that if the rules (causality and factivity) cannot apply, well, they cannot apply, and the rules should change. So that the empirical research on episodic memory, by showing that both the factivity and the causal conditions are impossible, strictly speaking, to fulfil, leads us to revise the epistemic norms of both remembering and successful remembering, and accordingly the functions of episodic memory. I contend that available characterizations of successful remembering are too demanding. Eventually I propose an explanation for this: the epistemic regime is not the only normative regime surrounding episodic memory. We do not evaluate memories in every context in function of their epistemic value. Epistemic evaluations are valid in a court of law, for instance. But we're not always trying to get the past right. In the following sections, I make a case for non-epistemic memory evaluations: evaluations of psychological benefits and potentiality. This leads us to revise the presupposed function of episodic memory. I contend that episodic memory is not only directed towards the knowledge of the past (section 4).

Section 1 – Epistemic memory and empirical memory as distinct

Memory science and the rediscovery of a constructive dimension of episodic memory puts pressure on the causal claim. Psychology of memory has recently put pressure on first-hand experience as a valid source of knowledge of the past (see chapter 3). Eventually, all those empirical works tend to bring memory and imagination closer, and this leads to a position called continuism (Perrin, 2016).

Continuism is for instance embraced by Michaelian, 2016a. Against this view, people have recently argued that the normative project and the descriptive or empirical project should be distinguished.

My aim in this chapter is to propose some limits to the objection. Though I also question Michaelian's 2016a, 2016c view, from the motive that it still considers memory as primarily directed towards the knowledge of the past, I contend that what the empirical research on memory leads to is actually to revise our presuppositions on the functions of episodic memory. I contend that memory is not always for the knowledge of the past, and that the episodic memory system has important non-epistemic functions. I will propose considering a further normative project which is overlooked too in the contemporary research on episodic memory.

In this section I start with presenting three ways this view has been defended in recent literature: the normative project is about the causal condition (Mahr and Csibra, 2018); first-hand experience has a specific epistemic value and plays a major role in memory evaluations (Craver 2020; Henry and Craver 2018); there is a fundamental normative discontinuity between memory and imagination related to their epistemic value (Barner draft).

1.1. Causal relations

According to Mahr and Csibra 2018, the research on the constructive dimension of episodic memory is related to a descriptive project within psychology that describes what memory actually does. But this should not compete with what memory is supposed to do. In two regards; first, episodic memories are presented to the subject as being caused by the past experience they aim at representing. The memory output seems like it is caused by an input, which is the past experience. Second, there are two different ways of asking the question of the nature of memory. One is a descriptive question and is focused on the best description of episodic memory. Another one is a normative question and is focused on the conditions that should justify considering memory as memory. The underlying claim is thus: if one

understands the causal theory of memory as engaged in a normative project and not a descriptive one, then the empirical arguments against it cannot hold.

About the first claim, one could object that it is unclear that memories present themselves as caused by the past experience. It only seems that they are about the past. An experimental philosophy study by [Dranseika 2020](#), goes this way. People are willing to call memories either genuine memories, false memories, or quasi memories. This suggests that people experience their memories as being about the past, but not necessarily caused by the past experience. In a sense, people know that memories can be caused by sources other than the very past experience their memories aim at representing. In a sense, thus, the first claim is too strong.

The second claim is about the possible or impossible interactions between the empirical and the normative projects about episodic memory, and poses a highly interesting question: what are the conditions that justify calling a memory a memory? Does memory science tell us anything about how memories should be evaluated?

1.2. First-hand experience.

This second claim is also endorsed by [Craver 2020](#); [Henry and Craver 2018](#). In his 2020 paper, Craver makes an analogy between episodic memory and the game of chess. In the game of chess, he argues, there are two distinct regimes: one is the normative, the one of the rules, that ‘describes proper chess play’ (Craver, 2020, p. 5); another one is the empirical, and is about all the ways chess can be played on a board. They have different success conditions and thus are distinct. Moreover, if empirical chess cannot happen without normative chess, the reverse is not true. Empirical chess does not change the rules of chess. Craver’s claim about episodic memory is built on this case. Normative memory is presupposed by empirical memory. Empirical memory depends on normative memory but without reciprocity. The success conditions of normative remembering do not depend on empirical remembering. Normative remembering is about epistemic success conditions; the number one rule of memory is that it is a first-hand experience (Craver, 2020), and hence it gives a privileged epistemic authority to the rememberer (Robins

forthcoming). If research on empirical remembering shows that memories can have sources other than first-hand experience, that does not change the fact that memory evaluation depends on this. Remembering is holding a ‘trump card’ (Henry and Craver, 2018), because it gives you an epistemic privilege of knowing the past thanks to first-hand experience. First-hand experience has a high epistemic value. The distinction between epistemic and empirical remembering is useful and clarifies a lot of things in the debate about memory accuracy. But, I contend, it faces two issues. One is that even in chess when some rules cannot be fulfilled in empirical chess, the normative chess adapts, and other rules apply. In specific empirical contexts, when checkmate is impossible, another rule applies: stalemate. So, the first issue faced by this view is to presuppose that the norms surrounding memory or chess are unified, that the same unique rule applies in every context. Second, whereas chess is a terminal game, of which we know the rules, memory is not. And Craver’s claim happens in a specific intellectual context which is precisely about determining the rules of memory. According to him memory is about ‘the game of knowing the past’, but this might not be the whole picture.

1.3. Normative discontinuism

The third important way of defending this claim has been proposed by Barner (draft) and focuses on normative discontinuism. In her paper, she shows that even though the empirical research on mental time travel might lead one to endorse continuism (Perrin and Michaelian, 2017; Michaelian, 2016a), and consider past oriented mental time travel (PMTT) and future oriented mental time travel (FMTT) as being of the same kind, there is a fundamental normative discontinuity between remembering the past and imagining the future. We do not expect the same kind of contents from those two capacities. Memories and future thoughts cannot be epistemically evaluated in a similar way. Memories can be evaluated in function of their epistemic value about the past, but we cannot expect the same from future thoughts. Memory is factive, imagination is not, because memory is about things that happened in the past and imaginations are about things that have not happened yet. We evaluate memories in function of their truth with regards to the past; the same rule cannot apply to imaginations of the future. This is a strong point against

continuism. But beyond the concerns I formulated against the factivity condition (Chapter 3), I also think that this again hides the fact that episodic memory is not all about the truth of the past. Episodic memory does not have only epistemic functions. This is what I intend to show in the following sections.

Section 2 – Epistemic evaluations of episodic memory depend on contexts

My first argument against the epistemic normative view is that epistemic norms of memory evaluations are not unilateral. There are cases where both the factivity and the causal condition cannot apply, because they cannot be fulfilled. And under those circumstances, other norms apply to determine the epistemic value of memories. Here, I build the argument on the claim that clinical memory distortions can have epistemic benefits when they are epistemically innocent, namely when one could not form a more accurate picture of the past, as proposed by [Bortolotti 2015](#); [Bortolotti and Sullivan-Bissett 2018](#).

2.1. No alternative condition and epistemic innocence

To go back to the chess analogy proposed by Craver 2020, it is unclear that the norms are always the same in empirical chess. On the contrary, the norms of chess cover a variety of possible situations and, under specific circumstances where some rules are unfulfillable, other rules may apply. For instance, in chess, there is a case where the checkmate rule cannot be fulfilled, and another rule called stalemate applies in its place. Thus, chess does not have a unique rule to determine how the game ends. Similarly, one could argue that in the memory case, under certain conditions, in a certain context, when the usual norms of epistemic evaluations are impossible to fulfil, other rules apply. And this might be the case in clinical memory distortions and confabulations. It is clear that clinical confabulations and memory distortions have high epistemic costs (these have been highlighted in particular by [Bortolotti & Sullivan-Bissett, 2018](#)), but does that mean that they have no epistemic

value at all? Confabulation is defined as the production of distorted memories about one's own past and the world without the intention to deceive, and as the attempt to fill in memory gaps, either spontaneous or motivated, either accurate or inaccurate.

It can appear in amnesia (Moscovitch 1995), but also in delusional states such as Capgras syndrome (Capgras and Reboul-Lachaux 1994; Frazer and Roberts 1994; Hirstein and Ramachandran 1997; Ramachandran 1998), somatoparaphrenia (Shallice 1997), and mirrored-self misidentification (Breen, Caine, and Coltheart 2000); in hypnosis (Coltheart 2017); and in a health context (Wilson and Nisbett 1978).

Though confabulation has often been characterized as a dysfunction (see Robins, 2016a), it has also been proposed that it is not dysfunctional in itself, as the phenomenon of confabulation could be seen not as an undesirable effect of but as a normal response to amnesia (Coltheart 2017; Gopnik 2000; Kopelman 1987). Gopnik, 2000, for instance proposes an evolutionary account of 'the drive for causal understanding' (Gopnik 2000, p. 311), which makes it possible to say that a bad explanation or a pseudo-explanation are not the same as no explanation at all), or as the consequence of a property of human cognition led by the drive for causal understanding (see Coltheart, 2017).

Bortolotti & Sullivan Bisset, 2018 have proposed two necessary and sufficient conditions for the epistemic innocence of clinical memory distortions. The first one is the no alternative condition, namely, the absence of an alternate explanation of or cognitive access to the past available to the subject. The second one depends on the epistemic benefits of the distorted representation of the past.

2.2. Epistemic benefits

Clinical memory distortions, despite their epistemic costs, might have epistemic benefits. When people construct a representation of a past event, even though inaccurate, they train epistemic skills such as belief formation, inferential reasoning,

consistency making, explanatory skills and construction of the representation of oneself. Confabulating could help in retaining key beliefs about one's self (Bortolotti 2015; Bortolotti and Sullivan-Bissett 2018; Letheby 2016; Sullivan-Bissett 2015). Bortolotti & Sullivan-Bissett, 2018 have thus argued that when distorted memories are the only available representations of one's own past, they have important epistemic benefits, and can be said to be epistemically innocent, because the epistemic benefits they have cannot be obtained otherwise. They are beneficial in at least two senses: they prevent harmful lack of self-representation and self-confidence, and they are a way of training important epistemic skills.

Moreover, constructing a representation of a past event even if inaccurate could be seen as a way of endorsing a position, that may allow people to exchange information and receive feedback on it (Sullivan-Bissett, 2015). Confabulating to construct reasons may be a way of becoming aware of my actions, attitudes, values, or beliefs. Seen from this perspective, confabulating could be seen as a way of entering the 'shared-narrative space' (MacIntyre 1981), where people can form and maintain a sense of self, ask and give reasons for what they have done, discuss their memories, receive feedback on them, and revise them. Constructing a coherent narrative of myself, regardless of its accuracy, could also be seen as a way of training epistemic skills and entering into epistemic practices of memory sharing and revision.

If in specific cases, accuracy is not what matters to epistemically evaluate memories, it seems that the epistemic norms surrounding episodic memory are less strict than presupposed by the normative epistemic view. On the contrary, in specific empirical conditions, when memory cannot deliver truths about the past, the norms of the epistemic value of memory change. It is thus unclear that the empirical and the normative regimes about episodic memory do not interact. The normative one depends on the empirical contexts where memories are evaluated.

Section 3 – Memory science leads to revising epistemic memory evaluations.

In this section I move towards a more radical view and propose that if the rules (causality and factivity) cannot apply, well, they cannot apply, and the rules should change. So that the empirical research on episodic memory, by showing that both factivity and causal conditions are almost impossible, strictly speaking, to fulfil (Chapter 3), leads us to revise more radically the epistemic norms of remembering, and to understand how to epistemically evaluate memories.

According to Craver 2020, epistemic remembering does not depend on empirical memory. I think this claim faces an important issue, which is the following: if the normative rules of memory evaluation cannot apply practically, in the empirical world, well, they cannot apply. If the norms of chess were unachievable in the empirical world, they would make no sense. What I contend here, is that the empirical research on memory actually should change our memory evaluations. If we are looking for knowledge of the past, episodic memory as an individual capacity cannot be enough. I have raised concerns against both the factivity and the causal condition in Chapter 3. It seems that those two conditions are practically impossible to fulfil. Even though they are presupposed by empirical memory science (Craver, 2020), research on episodic memory leads us to question those presuppositions. Episodic memory alone is not enough in ‘the game of knowing the past’. And this game, contrary to chess, which is a finite game, is still in the conception phase, in the sense that the question which is asked by philosophy of memory is precisely the question of the norms that should be chosen to evaluate memories and define episodic memory.

3.1. Clinical memory distortions and non-clinical memory distortions.

Here, I follow a path proposed by Forest and Gérardin-Laverge 2019 according to whom memory science leads to reducing the distinction between clinical confabulations and non-clinical memory distortions. What we learn from memory

science is that recollection is more likely to be an act of reconstruction of disaggregated pieces of information encoded in the neural connections involved during perception of an event (De Brigard, 2014). Episodic memory is thus a capacity that generates coherence from disparate pieces of information. Thus, episodic recollection appears also, partly, to be a capacity of filling memory gaps (De Brigard 2014; McClelland 1995). Consequently, episodic memory recall seems to share some features with confabulation. As episodic recall appears as a result of a complex mechanism of reconstruction, it makes no sense to say that it does not meet the no alternative condition before being contradicted.

Research on the cognitive benefits of false memories has shown that memory distortions could be beneficial for cognitive tasks. For instance, Howe et al. 2011 have shown that memory distortions and false memories (in healthy situations), in list retention tasks in particular, can enhance both insight-based problem-solving performance and speed. De Brigard 2014 has proposed that low fidelity in memory could also have benefits for creativity and argued that the production of false memories is correlated with the enhancement of creative processes (such as convergent thinking; see Dewhurst et al. 2011). McCarroll and Sutton 2017 have proposed that remembering from an observer perspective, and the switching of perspectives in recollection, could be seen as complementary means for our understanding of the past, thus giving an epistemic value to the observer-perspective memories, as they seem to engage in epistemic practices (understanding in their vocabulary) despite their (supposed) distorted aspect. In this regard, distorted memories, confabulations, and other types of memory errors could be considered as complementary means of making sense of reality. Even if not totally accurate, constructing coherent representations of one's own past could be a valuable way of exploring it individually and collectively.

In the end, accuracy in the details and the correspondence of a memory with regard to the past experience it represents might not be necessary for giving an epistemic value to a memory. Even in the cases where coherence outweighs correspondence, an episodic memory may nonetheless exercise a rememberer's epistemic skills. Constructing a memory, even an inaccurate one, is a way of confronting it, receiving feedback on it, revising it and adapting to a social and environmental

context. If, before revision or confirmation, an episodic memory cannot be taken as a necessarily accurate representation of the past and a source of self-knowledge, it nonetheless is a way of training epistemic skills and engaging in epistemic practices and understanding of one's own self.

3.2. Interactions between memory science and memory evaluations

Second, I don't think that the descriptive project and the normative project do not interact. On the contrary, I argue that acknowledging the constructive dimension of episodic memory is useful for memory evaluation. First-hand information, accuracy, factivity are not norms imposed by memory but by a specific social context. Memory accuracy and truth matters in the legal sphere. At the game of knowing the past, memory science shows that episodic memory might not be enough. To evaluate memories epistemically one should take into account their epistemic contexts of construction.

Craver's objection to the elimination of epistemic memory in favor of empirical memory is in particular directed towards Michaelian's 2016c simulationist theory. On this theory, memory science, by showing that episodic memory is constructive, leads to abandoning the causal theory of memory and to adopting a reliabilist theory of remembering. Episodic memory would be part of a simulational memory system that does need appropriate causation and memory traces but that 'functions, as it was meant to, to give us knowledge of the personal past' (Michaelian 2016c, p. 239). If I don't think that reliabilism is satisfactory, I nonetheless contend that empirical memory has something to say to the epistemic domain. Simulationism faces several issues though: one is that it has difficulties in explaining forgetting as a normal and common phenomenon (McCarroll forthcoming); another one is that it has difficulties in distinguishing between misremembering and confabulation (Robins 2017; Robins 2016a) for an answer, see Michaelian 2018. Another issue is that simulationism, since it is part of an internalist epistemic project, faces difficulties in explaining how memory could be a source of knowledge of the past if the outputs are not necessarily caused by past experiences. In a sense, how episodic memory can get the past right seems magical, if successful remembering

is not necessarily caused by past experiences. It also considers confabulations as produced by a dysfunctional system, and this is not so clear. For two reasons. First, as shown in Section 2, confabulations could be regarded as functional in certain aspects. Second, people have argued that non-dysfunctional cognitive systems could produce confabulations (on non-clinical confabulations, see [Wilson and Nisbett 1978](#), see also [Robins forthcoming](#)). Eventually, and this is the argument I develop here, simulationism, like most epistemic accounts of remembering, focuses on the differences between successful remembering and memory distortions and errors. In doing so, it undermines the features they share.

I argue that what memory science and the rediscovery of a constructive dimension of episodic memory could lead to is a more radical view regarding memory as an epistemic source. It is likely that episodic memories cannot be accurate pictures of past events and experiences (Chapter 3). They are produced in specific epistemic contexts where the background beliefs (regarding what is true, what happened in the past, and who is a trustworthy source of information) of the rememberer play an important role (Chapter 3 again). In this regard their epistemic evaluation should take into account their epistemic context of formation and construction. Since we learn that memories are labile and that their construction depends on the contexts in which the subjects encode and remember, then, how memories represent the past depends on their construction history.

3.3. How to know the past on the basis of memory reports? Information and interpretation

I thus contend that memory science, or the descriptive domain, leads us to revise the epistemic norms of successful remembering. The available characterizations of successful remembering put the stress on how memory allows us to represent the past accurately. Unsuccessful remembering on the other hand is unsuccessful because it does not represent the past accurately (Robins, 2016a). And confabulation is from another type; it is produced by a dysfunctional system, as such confabulations will generally be inaccurate (Michaelian, 2018). However, what memory science has shown for the past few decades is that memories barely

ever represent the past totally accurately. And, in practice, memories are not taken as sources of incontestable knowledge. In a lot of psychological experiments, subjects are asked to rate the confidence they have in their memories. The fact that people can rate several personal memories differently is a first evidence for the claim that they do not take all of their memories as accurate pictures of past events. When I remember, if my aim is to know the past, I am likely to consider several factors in my memory evaluation. Is the representation of the past I have likely? Is it coherent with what I know? Am I good at memory accuracy? What are my background beliefs? Similarly, when someone tells me a memory, if my aim is to know the past, and evaluate it in function of how it represents the past accurately, I am likely to consider several factors: is it coherent with what I know? Do I trust the rememberer? Do I trust the rememberer's capacity to remember accurately the past? What are the intentions of the rememberer? Is she joking? Is she serious? What are her background beliefs? What is her emotional situation?

In a sense, an issue that faces both simulationism and the causal theory of remembering is that, even though they are trying to give conditions for remembering the past accurately, their epistemic accounts cannot help to determine in practice if a memory is accurate or not, and in particular in situations where there are epistemic conflicts. Here is a puzzle for both the causal and the simulationist theories of remembering. Let's imagine that Paul and Marie lived a common past event: they skyped with their friend Constant. The next day, Paul and Marie both have a memory of the event and share it. Each of their memories could be described as caused by the past experience or produced by a reliable simulational system. But here is the problem. They disagree about what happened. Paul remembers a pleasant conversation where the speech was equally shared. Marie remembers an awful conversation where she had no talking space. They might enter into a discussion about what happened and enhance the accuracy of their respective memories in a collaborative enterprise (as has been described by [Harris et al. 2019](#) for instance). But let's imagine they don't. They are in conflict and don't trust each other enough to collaborate. They are both stuck in their perspective. Neither the simulationist nor the causalist can decide who has the most accurate representation of the past.

Maybe neither of them has a more accurate representation of the past, but in any case, in the absence of a recording of the conversation, to evaluate the memory accuracy in this case one will have to take into account the epistemic contexts in which both memories have been formed. If one finds out that Paul is sexist and often considers it as normal that women speak less, Paul's memory construction is likely to be dependent on those background beliefs. Marie's memory will be more likely to be more accurate. If one finds out that Marie has serious self-centeredness issues and that she often considers that she is not being given enough space, her memory construction is likely to be dependent on this background belief. Paul's memory might be more accurate. My point here is that in many cases memory evaluations cannot be done without taking into account the epistemic context of memory construction. If we want to determine the degree of accuracy of memories, we might need to study their construction history.

Moreover, it is uncertain that there is such a big distinction between successful remembering, misremembering and confabulation. What I want to argue for here, is not only that confabulation can have epistemic benefits related to the training of one's epistemic skills, but also that some confabulations can have an epistemic value in their content. I believe that confabulations can also be a way to give valuable information on oneself.

A distinction is needed here. There are different types of confabulations. Some appear as the result of error-targeting questions; they are motivated or provoked in conversation. Those are the main type of confabulation used by people focusing on the differences between confabulation and successful remembering. But there is another type. Some confabulations are said to be spontaneous. In spontaneous confabulations, people have been described as making errors in the recognition of the temporal order of stored information, but they generally give accurate information about the past. They only misplace events in time (Schnider, von Daniken, and Gutbrod 1996). In many cases spontaneous confabulations do give information on one's own history, interests, worries.

Thus, confabulations can be ways of communicating about myself. Then, if one wants to know my past on the basis of my confabulation, one will have to interpret

it. But my point is that this interpretative attitude might be needed in an epistemic evaluation project for both successful remembering and cases of confabulation.

Eventually, the empirical domain can give guidance to the normative one. If one wants to know the past on the basis on memory reports, one will have to take into account the contexts of memory formation and construction.

Section 4 – Non-epistemic functions of episodic memory

My final point is that the epistemic regime is not the only normative regime surrounding episodic memory. We do not evaluate memories in every context in function of their epistemic value. Philosophy of memory, by focusing on how episodic memory can be a source of knowledge of the past, might miss other important functions of episodic memory. Epistemic evaluations are valid in the court of law for instance. But we're not always trying to get the past right. In the following section, I explore some non-epistemic memory evaluations: psychological benefits and agency or potentiality evaluations. I contend that this leads us to revise the presupposed function of episodic memory and, I argue, to consider that episodic memory is not only directed towards the knowledge of the past. Moreover, both continuists and discontinuists focus on the epistemic dimension of episodic memory. I think they miss something. There are other rules surrounding episodic memory. And in particular, memories are not always evaluated in function of their epistemic value. There are non-epistemic memory evaluations, and there are non-epistemic memory functions.

4.1. Contextualism about function attribution

First, function attribution depends on the explanatory project one is engaged in (Nanay 2012). The epistemic regime is not the only normative regime surrounding episodic memory. We do not evaluate memories in every context in function of their

epistemic value. Epistemic evaluations are valid in the court of law for instance. But we're not always trying to get the past right. The idea that the function of episodic memory is epistemic is true in a specific explanatory project, when one wants to explain how memory can be a source of knowledge of the past. My claim is that the normative views presuppose the function they attribute to episodic memory, as being its true and primary function. But they are engaged in a specific explanatory project. If one's explanatory project is to distinguish between remembering and other types of mental actions, the idea that the specific function of episodic memory is epistemic makes sense. But if one's explanatory project is to determine the variety of things that episodic memory allows a subject to do, then the idea that episodic memory's function is to know the past might be insufficient. In the project I am interested in, namely, explaining what kind of role memory plays in the constitution of personal identity, the story might be different.

4.2. Beneficial vs/ Unbeneficial

4.2.1. Unbeneficial accurate memories.

First, there are contexts where memories are evaluated in function of their benefits for the well-being of the rememberer. Posttraumatic stress disorder (PTSD), for instance, is a syndrome where someone's memory, regardless of its accuracy, is characterized as unbeneficial (DSM-III, 1980) and harmful (O'Donnell et al. 2003; Vieweg et al. 2006).

PTSD syndrome is related to the retrieval of a traumatic memory of a past experience, though it is judged as unbeneficial. Therapy to overcome PTSD syndrome is focused on the possible ways of either making the subject forget the memory (Giustino, Fitzgerald, and Maren 2016) or distorting it and changing the emotional valence associated with the memory (see Kida 2019 on fear extinction and memory reconsolidation in PTSD therapy).

4.2.2. Beneficial memory distortions and psychological benefits of clinical confabulations

On the other hand, memories regardless of their inaccuracy can show benefits for the well-being of the rememberer and have psychological benefits. Some of the literature on confabulation puts the stress on this. It is first worth noting that attempts to challenge clinical spontaneous confabulations are often unsuccessful. Indeed, resistance to challenge is generally considered to be a characteristic of clinical confabulations and has been characterized also as sometimes non-beneficial (Zangrilli et al. 2014).

It has been shown that clinical confabulations, despite their epistemic costs, have psychological benefits (Bortolotti & Sullivan-Bissett 2018). In particular, confabulation might result in enhanced self-confidence and sense of competence as it would prevent both dumbfounding (Bortolotti and Sullivan-Bissett 2018; Kopelman 2010) and admitting ignorance (Hirstein 2005), sense of coherence (see Moscovitch 1995, where patient HW is described -- cited both in Bortolotti and Sullivan-Bissett 2018 and Robins 2017), and psychological well-being (Bortolotti, 2015; Bortolotti & Sullivan Bisset, 2018).

4.3. Inhibitory vs/ Empowering

A second non-epistemic way according to which some memories are evaluated depends on their effect regarding action potentialities. Here are two important memories of my childhood years.

Memory 1:

One day, I was with friends at a school party. We were playing around the skate ramp. Two friends jumped from the maybe 3 meters high platform. And then, they were saying to my sister, 'go ahead Mona, jump!'. She was hesitant. I don't know

whether this was out of pride or sacrifice, but I jumped, out of the blue. And I remember crashing, hurting my back really hard, and almost crying on the ground. That was a failure.

Memory 2:

Another day, it was during the summer, I was with friends at the swimming pool of my village. We were playing there almost every day during the summer break. One of the games was a competition. To swim all the 25 meters long under the water. We were all making it with our young lungs. So, one day we competed to see who could make it both back and forth. And I did a complete back and forth. That was a success.

The failure memory prevented me on many occasions from showing off, jumping from heights and things like that. This produced a sense of danger in these situations and a long-lasting vertigo. The success memory, though, produced and still is one of the bases of my idea of myself as being a good swimmer and having a deep breathing capacity.

One is at the basis of my sense of incompetence for specific tasks, one is at the basis of my sense of self competence for other kinds of tasks. The question of whether they represent the past accurately is secondary for them to play these roles. I can evaluate them despite their epistemic value. One of them has an agency value; one of them has an inhibitory value. Of course, they are both constitutive because I believe they come from my own experience. But the question of whether they are true or not does not matter for them to play a constitutive role.

My point here is that there are different ways of evaluating those two memories. One could evaluate them in function of their epistemic value. But this is not the whole story. I argue on the contrary that what matters often in memories is what they carry as representations of oneself, and as such what they offer as potentialities of action. This might reveal an overlooked role of episodic memory in contemporary philosophy of memory: memory is not only used to know the past; it is also used for self-constitution and action. Such an action-related functional account of episodic memory has been studied in psychology by Maguire and

colleagues, for instance. Episodic memory has an important function related to action-planning (Maguire et al., 2000; Hassabis et al., 2007; Maguire et al., 2010). I will explore this topic further in the next chapter.

Section 5 – Concluding remarks

In this chapter I have discussed the view according to which empirical memory science is distinct from the normative domain. I have shown that epistemic memory evaluations should depend on contexts. I have first proposed that the epistemic norms are not the same in clinical and non-clinical contexts. More radically, I have argued further that memory science leads us to change our criteria for determining the epistemic value of episodic memories. I have proposed that taking into account the constructive dimension of episodic memory could enhance practically our normative evaluations of memories. Eventually I have proposed that as there are non-epistemic memory evaluations there are non-epistemic memory roles. If episodic memory is mostly regarded by contemporary philosophy of memory as a system dedicated to the knowledge of the past, that is not the whole picture. One possible objection to this claim is that the other roles I have proposed depend on memory accuracy, and memory evaluation as a source of accurate representation of the past. A possible answer to this objection could be that, as I have shown in Chapter 1, auto-noetic consciousness can happen in spite of memory inaccuracy. I can believe something happened to me even though it is a *false* memory. Thus, some memories can be constitutive of my personal identity regardless of their accuracy. Their psychological benefits and the potentialities they open to me as an agent might depend on the belief I have that they happened to me, namely on the auto-noetic consciousness associated with them, but do not depend on their accuracy. Episodic memory is not only dedicated to the true knowledge of the past, it serves other important purposes such as self-constitution, psychological well-being and action related goals. These have different success conditions, that do not necessarily involve memory accuracy with regards to the past events or the past experiences.

Third Part
– Rethinking memory to redefine
personal identity.

Chapter 5

– Personal identity, constructive memory and narrativity.

Introduction

Remind me who I am, again! is the title of the novel that Linda Grant (Grant 2011) dedicates to her mother, who was suffering from a memory pathology. The theme of the one who, losing her memory, loses her relation to herself reminds us how much our identity is carried by our memory. The question of the relationship between memory and personal identity has been an essential question in the history of philosophy since the classical age, but it is also crucial for contemporary societies faced with pathologies such as Alzheimer's disease. Not only is it widely believed that memory allows a person to recognize herself as oneself, but one of the consequences of many memory pathologies is to disturb the sense of identity without which it becomes difficult to act. The first part of this thesis has followed a tradition that privileges memory as a constitutive dimension of personal identity, but it seemed that the concept of memory itself needed to be revised to play this role. I have attempted to revise the concept of memory, grounding on contemporary research in the field of memory sciences, not with the intention of dissolving philosophy into neuropsychology but, from a philosophical perspective, to establish a dialogue between disciplines that seems called for by the subject itself and the state of available knowledge. In part 2, I have attempted to construct a concept of

memory consistent with empirical research. I have questioned the necessity of the factivity condition and the causal condition for remembering. I have shown that episodic memory has a double dimension: at encoding, it is a capacity that preserves traces of past experiences, by selecting features in the past experience. Retrieval consists in the construction of a representation of the past on the basis of those preserved traces and depending on the context of retrieval, that both triggers the memory traces, and orientates the construction. I have also questioned the functions of episodic memory and argued that the epistemic role was not the only role of episodic memory. I have proposed that rememberers use episodic memory to know the past but also to produce a representation of themselves and their past that is related to their potentialities for action. In this third part of the thesis I intend to go back to the question of personal identity. My starting point was an exploration of a well-known memory theory of personal identity, Locke's. The challenge of this last part and present chapter is to revise the philosophical concept of self. It is thus, on the one hand, a contribution to the philosophical debate on the role of memory in the constitution of personal identity. On the other hand, because of the proliferation of memory pathologies that alter the sense of their own identity in amnesic patients, it aims at proposing a conception of the self that could be a useful contribution, in the clinical and medical field, to the accompaniment and care of these people.

The question I intend to ask again is how can a self extend itself through time despite the changes that affect it; and ultimately, what can be a constructive memory theory of personal identity? The picture of episodic memory that has emerged from empirical research, as I have shown in previous chapters, is the picture of a constructive capacity, dependent in particular on the context of construction. One important context is the social context in which memories are formed and retrieved. In this regard, a revised theory of personal identity might look very much like a narrative theory, and in particular MacIntyre's, 1981. According to MacIntyre, to be a person is to be able to produce self-narratives. Self-narratives are embedded in a shared narrative space, where they play a role in action explanation. In this last part and chapter, I explore the possible reconciliation of a memory theory and a narrative theory of personal identity, from the revised concept of episodic memory. I start with a review of the missing pieces of a memory theory in light of what episodic memory actually does (Section1). I go on and present MacIntyre's theory

as a potential solution (Section 2). However, after discussing the common presuppositions of Locke's and MacIntyre's theories (Section 3), I show that MacIntyre's theory is not immune to objections either. In particular, I relate Strawson's arguments against narrativity, and show that MacIntyre's theory is too vague on the tools of situated narration (Section 4). I nonetheless propose that some of his concepts are useful for an understanding of personal identity, but I adapt them to a memory framework (Section 5). Eventually I make an attempt to redefine selves in light of a revised concept of episodic memory. I propose that selves are individuals that extend through time thanks to a constitutive epistemic relation to themselves, and that constructive memory is a capacity that can play this role. It allows one to construct representations of oneself, and as such, to it leads self-recognition and self-constitution. As I have shown that factivity is not necessary for memory, I argue that confabulations also can be the basis for self-constitution. I eventually consider the question of amnesic selves.

Section 1 – The missing pieces of memory theories

1.1. Construction

Philosophical research on the relationship between memory and personal identity is part of a lively debate within modern philosophy on the definition of the individual. Memory theories of the self have their source in particular in the philosophy of John Locke. In *An Essay concerning Human Understanding* (1694), Locke proposes a psychological reductionism, since he reduces the self to the same consciousness (i.e. the same self-consciousness or the extended diachronic self-consciousness of the person's present and past). His theory of personal identity has the merit of answering both the epistemic question of personal identity (What makes me know that I am myself?) and the metaphysical question (What defines personal identity?). However, if we limit ourselves to the Anglo-Saxon field, two types of objections traditionally put it at risk. One denounces its circularity;

formulated by Joseph Butler (1736), it shows that consciousness presupposes identity and cannot therefore found it, since in order to be conscious of oneself, one must already be a self. Another one, that of Thomas Reid, is based on the weaknesses of memory: how can personal identity be based on memory if memory is a fallible faculty? I have discussed these objections in Part 1 and have shown that the concept of memory needed to be rethought. In particular, I have argued against Locke's preservative view of memory. Further, in part 2, building on what we know about episodic memory I have shown that memory has a constructive dimension. It is not only a preservative capacity. Thus, Locke's concept of memory misses its constructive dimension.

1.2. Social⁵⁶ dimension

Second, what can be criticized in the memory theory of personal identity is, if not its psychological reductionism, its ontological reductionism, which posits a person independent of a world and a context, an isolated person, alone with her consciousness; an ontological reductionism that forgets, in short, the social dimension of the self. But individuals evolve in a world, and individual memory has an intersubjective dimension (Gérardin-Laverge and Gérardin-Laverge 2015). Memories are formed and retrieved in a social world. On one hand, my memories are sensitive to the context in which they are formed and retrieved, as I have shown in Chapter 3; their construction depends on their formation contexts. On the other hand, rememberers are often not alone when they experience events, and when they remember them. In those cases, the same event can be remembered by more than one rememberer. When we target collectively a past event, adjusting our memories to retrieve the past is common. Either because rememberers have different representations of the past and look for what really happened. Or because one rememberer forgot parts of the event. This is the sense of a remark made by (Leibniz 1765) in his answer to Locke:

⁵⁶ I use 'social' to mean both interpersonal, and (perceptual and experiential) context-dependent. In this regard it is a weak sense of social, which does not involve a reflection on the role of institutions for instance.

“And if I were to forget all the past things, and be obliged to let myself be taught again up to my name and up to reading and writing, I could always learn from others about my past life in my previous state, as I kept my rights, without the need to divide myself into two people, and make myself heir to myself. And all this is enough to maintain the moral identity that makes the same person.”

Leibniz, New Essays, II, xxvii, 9.

Leibniz thus highlights the role of others in the constitution of my personal identity. There is a social dimension of remembering that is missing in Locke’s theory, but which is essential to memory processes, and can help maintain personal identity.

1.3. Envisioning one’s future.

Eventually, the matter of personal identity is not only related to self-extension to the past. One crucial question related to personal identity is: who will be me in the future? Again, this has been highlighted by Leibniz in the New Essays. He writes:

“An immaterial being or a spirit cannot be stripped of any perception of its past existence. It has impressions of everything that has happened to it in the past and even hints of everything that will happen to it.”

Leibniz, New Essays, II, xxvii, 14.

The relationship to one’s future is also an important topic of neo-Lockean reflection. (Parfit 1984; Velleman 1996) for instance have questioned the notion of psychological continuity not only between a past instance of myself and a present instance of myself, but also between my present self and a future instance of myself. There is an important literature on those questions, involving thought experiments such as the fission case (Lewis 1976). I will not go into the details of those here.

Nonetheless, it is important to mention that the relation to the future is missing in Locke's text.

However, some people have recently argued that episodic memory is crucial for the simulation of a personal future or for future oriented mental time travel. One important fact to support this claim is that people with episodic amnesia have difficulties both in past oriented mental time travel (PMTT) and in future oriented mental time travel (FMTT).

Tulving, 1985 is one of the first to give evidence for this. He notices that episodic amnesia, i.e., amnesia where patients lack detailed recent memories and have difficulty fixing new memories, also appears to result in difficulties in imagining the future in first person. In his paper, Tulving related the case of N.N., an amnesic patient. He shows that N. N.'s amnesia affects his personal memories and more generally his relationship to himself and to what he calls the subjective time. When asked what he did "yesterday", N.N. answers that he does not know, in the same way as when asked what he will do "tomorrow". Here is precisely what he says about the future:

"E. T.: Let's ask again about the future. What will you do tomorrow? (There is a 15-second pause)

N. N. smiles and says: I don't know.

E. T.: Do you remember the question?

N. N.: About what I will do tomorrow?

E. T.: Yes. How would you describe your state of mind when you try to think about it?

(A pause of 5 seconds)

N. N.: Blank, I guess" (Tulving, 1985).

So, N. N. not only has difficulties in remembering his personal past, it is also difficult for him to project himself into the future, to simulate it. The hypothesis that arises from these observations is that in episodic amnesia, it is not only the relationship to the past that is altered, but it is an episodic form of relationship to oneself, which extends from the past to the future. It is in fact the ability to mentally travel in time in first person that is jeopardized. The reflection on one's relationship to the future is also a missing ingredient of Locke's memory theory. But it is important for at least two reasons: psychological continuity depends on self-extension in a subjective time that comprises past and future. Second, it appears FMTT is not unrelated to PMTT.

I have thus shown in this section that Locke's memory theory is missing at least three important ingredients of what is needed for psychological continuity and also of what is needed for a description of episodic memory consistent with empirical research.

Section 2 – MacIntyre's narrative view.

Two aspects of the problem lead me now to explore MacIntyre's narrative view of personal identity. First, the missing pieces in Locke's theory are present in MacIntyre's. Second, if personal identity depends on memory and memory is a constructive capacity, depending on people's projection in the future, and with an important social dimension, it could be closer than expected to a narrative capacity⁵⁷.

⁵⁷ In this chapter, I focus on MacIntyre's narrative theory of personal identity, and neglect more recent versions of the narrative views, such as [Dennett 1992](#); [Hutto 2017](#); [Ricoeur 1992](#); [Schechtman 1996, 2001](#); [Taylor 1989](#); [Velleman 2005](#). It is possible that some arguments against the narrative view that I address in section 3 do not hold for all narrative views. However, I focus on MacIntyre's for at least two reasons: one is that he was the first one to propose such a view, and that his view is a direct answer to Locke's and what he calls the modern view of the self. Another one is that I am specifically interested in some of the concepts he proposed, such as the shared narrative space, the concept of embedment and the concept of narrative adjustment. I believe that they can be useful for a better understanding of both episodic memory construction and of personal identity recognition and constitution.

2.1. Construction of a coherent representation of myself as the foundation of personal identity

For MacIntyre, what grounds the continuity of a person is the belonging of the episodes of his or her life, and therefore of his or her individual acts, to a whole to which he or she can give reason. The unity of a life, at the foundation of personal identity, lies in the intelligibility of a person's actions, that is, in their coherence. According to MacIntyre, it is the narrative I can give of myself that makes any of my actions intelligible, and if the narrative can explain the sequence of my actions it is first because I live my life as a narrative, that is, I am inscribed in a lived narrative even before telling it. Personal identity is therefore ultimately based on the narrative(s) of oneself. But the narrative is not fixed. There are always several possible descriptions of a sequence of actions, and different versions can be given at different moments of time depending on what is to be explained more specifically. A narrative capacity is thus a constructive capacity. On the one hand, there is a multitude of possible narratives of the self, and the best narrative is constructed in dialogue; on the other hand, a self-narrative is always embedded in frames according to which the self-narrative takes shape.

2.2. Frames of self-narratives, and the concept of embedment. Individualistic worry.

My self-narrative is embedded in settings, namely other narratives such as the narratives of my family, of my time, etc. These frames have a story of their own. My self-narrative construction depends on the contexts in which it is formed. Self-narratives are embedded in an interlocking set of narratives which determine my own. My self-narrative is embedded in a historical context, in a social and an intersubjective context. Against the modern individualistic view of selves, MacIntyre insists on the frames that put constraints on my own life and on my way of making sense of it. He gives a criticism of Americans who think they have nothing to do with the history of slavery because they themselves never had slaves. According to him, denying where I come from, what is the history of my country,

community, etc. is a mistake. I am always dependent on this context, because I benefit or suffer from it, or in any case it has consequences on how I act, should act and who I am.

Beyond the narrative frame, and the settings in which my self-narrative is embedded (MacIntyre, 1981, p. 206), I am included in a social world. Others can supplement my self-narrative, as suggested by Leibniz (New Essays, II, xxvii), but can also produce a narrative of me different from mine, and thus can compel me to modify my self-narrative. According to MacIntyre this aspect is captured by the concept of a shared narrative space. When I produce a self-narrative, it enters a shared space where people can ask and give reasons for what they themselves and others have done. A self-narrative always is a matter of adjustments and revisions. Whoever produces a different narrative might have to justify it and require justification from myself if I want to hold to my version.

According to MacIntyre, being a person is thus being able to give a narrative account of one's actions. Such a theory entails that persons cannot be simply defined as entities to which we can ascribe both physical and mental properties. What matters is the ability to make our past actions intelligible to us and to others by a coherent narrative. In this regard, for MacIntyre self-knowledge should be considered as a "shared achievement". And with his idea of a shared narrative space, where people can give and ask for reasons for what they have done, MacIntyre invites us to think about the co-construction of self-narratives.

2.3. The meaningless life worry vs intelligibility – narratives and intentions

In *After Virtue*, MacIntyre has proposed a narrative view of personal identity which is an answer to several dissatisfactions in the philosophy of persons. And one of them concerns Lockean views of psychological continuity. In Locke, the self is constituted only by the set of episodes one can be conscious of and can attribute to oneself in a first-person perspective. Personal identity depends on consciousness as a self-consciousness and on the consciousness to be the same person. According to

MacIntyre, such a theory presupposes an anteriority of parts on the whole. To him psychological continuity of an individual cannot be enough to make a person.

To illustrate this worry, [Gérardin-Laverge and Forest 2018](#) have proposed the following thought experiment. Let's imagine a creature that remembers episodes of her past. But the episodes are not related to each other. The action performed in June does not extend or supplement any of her earlier actions, and she cannot explain it as part of a whole. No long-term goal or intention is available for considering several of her actions together. The possibility of linking episodes through memory seems thus an insufficient condition for personal identity, in the sense that it is not enough to be the author of several actions, if I merely see that I have done them.

According to MacIntyre though, for episodes to be co-personal, authorship or memory are not enough. They have to make sense in the continuity of my intentions. Such intentions can change, but for the past episodes to be intelligible to me and others and as such to make up my personal identity, they have to be explainable in a self-narrative. But not only that, they also have to be intelligible in the context in which they are performed. To illustrate this point, he proposes the case of a recipe in a cookery book. The actions described are intelligible in the context of a sequence that aims at baking a cake. Not only they are intelligible thanks to a narrative of how to bake a cake, but they also need a specific context to make sense. MacIntyre writes:

“If in the middle of a lecture on Kant’s ethics I suddenly broke six eggs into a bowl and added flour and sugar, proceeding all the while with my Kantian exegesis, I have not, simply in virtue of the fact that I was following a sequence (...) performed an intelligible action.” MacIntyre, 1981, p.209

The intelligibility of a narrative about one's actions is dependent on a context in which those actions are performed. What makes the continuity of a person's life is the belonging of her life episodes –and of her individual actions—to a whole that she can make intelligible. The unity of a life grounds the unity of a person, and depends on the intelligibility of a person's actions, namely of their consistence. And

what makes my actions intelligible is the narrative I can produce about them. The narrative is explanatory only because I live my life as a narrative and I am inscribed in a lived narrative before telling it as a story.

So, we see two absolutely antagonistic and apparently unreconciled positions emerging on the question of personal identity. On the one hand, the memory theory of personal identity, from its first version to its contemporary versions, places memory on the level of authenticity and objectivity, since memory, as long as it functions, is considered as a preservative capacity that records, encodes and recovers memories in an authentic way. It is the truth of the self. Within this framework, memory theory grounds the continuity of the person on the objectivity and authenticity of memory. Even in the case of a quasi-memory, if the memory is not a memory of my experience, it is indeed a memory that is faithful to a past experience and grounds psychological continuity. On the other hand, narrative theory tends to show that this continuity is subjective and constructed within a narrative that is embedded within other narratives.

Section 3 – Incompleteness of memory and narrative theories – their common presuppositions.

These two positions in their primary form may seem insufficient. First of all, memory theory seems to isolate the person without understanding her as a subject anchored in a context. It seems, moreover, that Locke's memory theory forgets a dimension of the person that is nevertheless fundamental: her relationship to the future. In fact, if memory is the foundation of the continuity of the person between past and present, we can wonder whether the capacity to project oneself into the future is not constitutive of the personal identity in time.

As for narrative theory, it does not explain what the tools of situated narration can be. In particular, MacIntyre never uses the notion of memory (without which it is difficult to understand how a person can be part of a history with legacies, traditions, etc.). It is unclear how without memory one could be able to give a self-narrative

related to her own history. Even though there can be several possible descriptions of a same past event, for those descriptions to be descriptions of this past event, one needs to remember it. Focusing on narrative construction thus undermines the importance of the preservative dimension of memory for maintaining a sense of self. Plus, he remains unclear about a person's relationship to the future (while he mentions the importance of longer-term goals for guiding actions, the relationship of these goals to a complex volitional structure remains unexplained). The radical alternative between the truth of memory and the constructivism of narrative is still awaiting conceptual clarification.

In fact, it still presupposes that memory is an activity of retention of information, and that it is above all individual: Forgetting is stigmatized as a simple dysfunction of a fallible faculty and we speak of "false memories", memories of events that we have not experienced on our own, but that we remember as episodes of our own life, that we integrate into our autobiographical memory spontaneously or after suggestion, for example (see [Loftus 2005](#), and the discussion in Chapter 3). However, as I have shown in part 2 of the thesis, the image of memory that emerges from scientific research is very different from its ordinary conception. And one of the challenges of the research on memory today is to invite us to revise our ideas about its mechanisms, functions, limits and integrity. If the function of memory is indeed to be a faculty of retention in the socio-legal framework (when a person is asked to testify, he or she is asked to swear to tell the whole truth), it is not certain that this is generally the case. On the contrary, I have argued in Chapter 4 that episodic memory has non epistemic functions.

Section 4 – Against narrativity

In this section I give reasons not to endorse a narrative theory of personal identity. I focus on Strawson's objection to narrative theories in general and to the ethical Narrativity thesis in particular. I argue that the fact that some human individuals do not have a narrative access to themselves is enough to deny the idea that narrativity is essential to their being as persons. It should prevent us from defining persons in

terms of narrativity. I propose two further cases as reasons to reject a narrative theory of the self. One is the case of nonbelieved memories. I argue that the idea of a coherent mental life, at the heart of a narrative theory, is incompatible with empirical data on beliefs. People hold contradictory beliefs about the past, and even remember nonbelieved memories. Another one is a case of psychological continuity without narrative capacity. I present the case of B, Storrington's patient, and argue that even though he cannot form a self-narrative, he is able to maintain a psychological continuity in time.

4.1. Episodic selves.

The idea that personal identity depends on a narrative account of oneself is popular. Beyond the version abovementioned, versions of the narrative view have been proposed for instance by [Dennett 1992](#); [Hutto 2017](#); [Ricoeur 1992](#); [Schechtman 1996, 2001](#); [Taylor 1989](#); [Velleman 2005](#), etc. It is also endorsed by some psychologists. [Merrill, Waters, and Fivush 2015](#) and [Waters and Fivush 2015](#), for instance have proposed that memory coherence is involved in maintaining a narrative identity and enhancing psychological identity. As described by Strawson, there are two main claims in most narrative theories (including MacIntyre's). One is that persons are storytellers, who live episodes of their lives in a narrative mode. Actions and lives have a basic narrative character, we live our lives as stories. This is what he calls the descriptive, empirical, or psychological Narrativity thesis. The second claim is that self-recognition depends on a narrative access to myself; being a person is being able to give a narrative account of one's actions. And to be a person and live a good life they must live their life in a narrative form. He calls this view the normative or ethical Narrativity thesis.

Strawson argues that those two claims should be rejected, the second claim in particular. His main argument is that there are non-narrative persons. He takes himself as one of them:

"I need to say more about the Episodic life, and since I find myself to be relatively Episodic, I'll use myself as an example. I have a past, like any human being, and I know perfectly well that

I have a past. I have a respectable amount of factual knowledge about it, and I also remember some of my past experiences 'from the inside', as philosophers say. And yet I have absolutely no sense of my life as a narrative with form, or indeed as a narrative without form. Absolutely none. Nor do I have any great or special interest in my past. Nor do I have a great deal of concern for my future.” (Strawson 2004, p. 433).

According to Strawson, the narrative theorists presuppose that their way of living their own lives is a universal way of living a life. However, this undermines the fact that not everyone lives their life as a narrative and that an individual can recognize herself through time without having to organize her past actions in a narrative form. Strawson argues that he has both semantic and episodic personal memory, though he does not care about his past. He argues that this does not prevent him from being a responsible individual, and a good friend. As an episodic self he is self-conscious, can recognize himself in time, has a sense of responsibility and is capable of commitments or at least involvement in long term relations. The non-narrative self might not be a person, but he has all the characteristics of a person. I take him as a counter example and a sufficient reason to deny that we should define personhood in terms of narrative capacities. However, one can argue that a narrative access to myself gives me a specific way of relating to myself and can help in recognizing me as the same person. Thus, narrativity should not give an answer to the metaphysical question of personal identity but could maybe give an incomplete answer to the epistemic question of self-recognition. Narrative capacities might help certain individuals to relate to themselves and to have a constitutive epistemic relation to themselves.

4.2. Coherence

Narrative theories focus on life-coherence. They describe self-access and self-constitution in terms of the construction of a coherent self-narrative. MacIntyre even proposes that life is a narrative quest with a unity, which is the unity of the

narrative. I think that a distinction is needed here. It is one thing to claim that when people explain their actions, they produce a narrative about causes and effects, give their reasons for acting etc. It is another thing to claim that people have a coherent view of themselves in every regard. There is a myth about coherence, according to which everything in subjects' mental lives is rational and is coherent, which might not fit to how we experience it.

On the contrary, it might seem that individuals hold contradictory beliefs all the time. At least it is a matter of debate. Psychologists have recently proposed that subjects could retrieve non believed memories (Clark et al. 2012; Mazzoni, Scoboria, and Harvey 2010). The idea is this: subjects episodically remember an event, though they believe that the event did not happen. The existence of unbelieved memories is a challenge for reflection on episodic memory. There is a debate on whether nonbelieved memories are plausible (see for instance Otgaar, Scoboria, and Mazzoni 2014; Otgaar, Scoboria, and Smeets 2012). Indeed, the idea that one has a memory but does not believe it happened contradicts the mainstream philosophical definition of episodic memories. Remembering needs to give rise to the idea that I lived the past experience I am remembering. It is supposed to be part of the phenomenology of remembering. If I don't believe that I experienced a past episode, I can no longer remember it. I can say that I remembered it but it was false, but I cannot say I remember going to the zoo but I don't believe this happened. Because remembering going to the zoo is believing it happened. So that in the case of unbelieved memories I would hold two contradictory beliefs, and this is implausible. However, if one accepts that a subject can hold contradictory beliefs, unbelieved memories could be more easily understood. I can remember going to the zoo last Thursday with Magdalini, because I have a mental image of going to the zoo last Thursday as though I had experienced it, but I believe that I did not go, because for instance I know that neither I nor Magdalini were available last Thursday, plus she told me that we never went to the zoo. I could revise my memory, and I do: I don't believe it happened anymore, though the memory persists. I think that if we abandon the myth of coherent mental lives, we can make better sense of these cases.

It seems that the myth of coherence is endorsed in a specific way by Schechtman, for instance. Being a person is not only to have a coherent self-narrative but also to have sympathy for every part of my personal past. On her account, for me to consider a past experience as mine, I should have empathy for it, a ‘deep sympathy’ (Schechtman 2001), or ‘a positive warm affective connection to the past events accessed in memory’ (McCarroll 2019, p. 261). Even though Schechtman mentions that it does not have to be positive all the time, there is something like a deep connection, a sympathy. To self-identify, or to recognize myself I should have sympathy for myself. It seems to me a very wise point of view. One that I am not capable of. In the sense that it could be seen as an account made by someone who does not feel retrospective shame, self-disgust or self-rejection, or even mere indifference for her past selves. It seems to me that I can self-identify with the remembered subject of some of my memories even though I don’t have sympathy for her, and I might be indifferent to myself in some cases. For instance, I did many things last week, like going to the grocery store, taking showers, having dinners; I have a lot of memories of last week, they are personal, they give me an idea of myself, and of what I did, but I have no specific feeling for their agent. I neither accept her nor reject her, it is just me in a notional sense (Velleman 1996). In the end, I am not sure that extending through time necessarily involves self-acceptance, sympathy for oneself, and other forms of ego trip.

I have discussed two ways narrative views focus on coherence; one is that a life is a coherent narrative, with coherence in all of its parts. I have argued that this might undermine the fact that we hold contradictory beliefs and unbelieved memories. I have thus proposed that even for one specific remembered sequence, for instance, one does not need to have a coherent version, in every regard. For instance, it might be possible to remember something and believe it did not happen. Another coherentist narrative view is Schechtman’s view of sympathy. Being a person and extending myself to the past supposes that I have sympathy for the subject of my memories. I have argued that there are cases on the contrary where I self-identify without having specific feelings towards my past self.

4.3. Personal identity without narrativity. Non-declarative episodic memory.

Second, it seems that people can have episodic memories and maintain a psychological continuity with themselves without being capable of forming a narrative including them. Though such memories have behavioral consequences. Such an example seems to be given by Craver 2012. Craver presents the case of B as a case of episodic amnesia; he uses this case to show that some people with episodic amnesia can maintain a sense of self, and an emotional continuity. This is drawn to refute a memory theory of personal identity, but it seems to me to be a case of episodic memory. In fact, it seems to me to be a case of episodic memory with a behavioral expression and no language expression. The case is thus: B reenters a room in which he had a traumatic experience, and even though he does not have a declarative memory of it, he runs away.

“At one point during his investigation of B, Störing returned with B to a hospital room in which B had received a highly invasive and aversive procedure to assess the shape of his cerebral ventricles. As soon as he re-entered the room, B became agitated and broke away from Störing. He opened the door, ran down the hall, and descended a flight of stairs. He even ran out the hospital before Störing could retrieve him. This fleeing action lasted longer than the two seconds of conscious awareness at B’s disposal, but the independent movements (opening the door, turning down the hall, leaving the building) were bound in a meaningful act by an emotional undercurrent carrying B from one moment to the next.” Craver, 2012, p. 457

Craver uses this story to make a point against the necessity of episodic memory to link experiences, claiming that emotions and mood can link conscious states at different times. Later in the paper, he writes, “one might be motivated by a fear that has lost its object in explicit memory” (Craver, 2012, p. 457). So, it seems to me that Craver’s point actually is not against memory here but against the idea that explicit memory is the only memory that links someone to her past experiences.

The behavior of B is best explainable by B's past experience, and for this past experience to have consequences in B's behavior, he has to remember it. So that B's case is not a case against the fact that memory is needed for psychological continuity but is really against the fact that explicit or declarative memory is needed. Nonetheless, the type of memory needed here is memory for episodes. The traumatic experience B had in this hospital room is a specific episode of his past. He did not learn to fear this room. He had a traumatic experience there once. So, in this regard, this fits with the concept of episodic memory, though he does not have an explicit memory of it. If episodic memory has been regarded as part of a declarative memory system, or part of explicit memory until now, it is unsure that our current categories are equipped to describe such a case. This case might be better described as a case of non-explicit episodic memory.

Therefore, this case seems to be evidence for the claim that episodic memory does not need to be declarative to play a role in persons' lives. And further, narrativity is not necessary either.

Section 5 – An attempt to revise the memory theory of personal identity: a constructive memory theory.

In section 4, I have considered the limits of narrative theories as theories of personal identity. Nonetheless I think that some of the tools of narrative theories are useful for understanding some aspects of memory construction. In particular, my main interest in narrative theories is that they allow us to think about how self-constitution depends on intentions and on a social context.

People have argued that the discovery of a constructive dimension of episodic memory should lead us to endorse a narrative view. One of Schechtman's motivations for endorsing a narrative account of the self, for instance, is that memory construction sometimes favors coherence over correspondence. On her view a revision of the concept of memory should lead memory theories of personal identity to endorse a narrative view (Schechtman 1994, 1996). [Schechtman 1994](#)

argues that memory summarizes and condenses past experiences and as such the memory view about personal identity, where memories provide simple connections to past experiences and between the rememberer and the remembered and thus are the basis of personal identity, faces issues. Instead, we should endorse a narrative view of personal identity where memory is a tool for the constitution of personal identity. I am sympathetic with this point. And I think Schechtman actually gives an answer to the worry on the missing tools of situated narration (Section 3) that I addressed to MacIntyre's theory. Schechtman's view has the merit of proposing a narrative view that considers the role and importance of memory and takes the constructive dimension of memory into account. But one could wonder why a narrative view is necessary if one adapts a memory theory to what we know about memory. In this section I give an attempt to revise a memory theory in such a way. I propose a constructive memory theory of personal identity.

5.1. Envisioning the future.

I have already mentioned Tulving's hypothesis on the relations between remembering the past and imagining the future. This has become a stronger hypothesis when confirmed by other investigative methods, including brain imaging studies. In 2007, Addis, Wong and Schacter, published a paper on common neural pathways of episodic recall and simulation of a personal future ([Addis et al. 2007](#)). In the experiment, participants are led to form a representation of a past or future episode over a given period of time starting with a given word. When they are ready, they press a button and still have twenty seconds to elaborate the event. It appears that the same regions are active in the construction of past or future scenarios. Again in 2007, Schacter, Addis and Buckner published another text in which they defend the idea that one of the crucial functions of the brain is to use stored information to imagine, simulate and predict future events. In this framework, memory appears as a tool of the brain not only to remember long-term intentions, but also to generate adequate simulations of future events ([Schacter, Addis, and Buckner 2007](#)).

Finally, the authors formulate the constructive episodic simulation hypothesis. They hypothesize that one of the adaptive functions of constructive episodic memory is to allow individuals to simulate or imagine future episodes. They rely on empirical evidence that shows an overlap between the psychological and neural processes involved in remembering the past and imagining the future. But if the hypothesis of constructive episodic simulation can shed light on the relationship between personal past and future⁵⁸, it is based on a leap between a proximity of neural activations and a causal role that episodic memory plays for future oriented mental time travel (FMTT). If the overlap shows that the activities may be of the same kind, it does not necessarily show their causal involvement, or their functions in relation to each other. That is, it does not tell much about whether the role of memory has anything to do with the imagination of the future. Indeed, it is one thing for PMTT and FMTT to share brain activity, it is another thing for memory to be constructive for the future. In Bartlett, for example, thinking of a constructive dimension of memory had little to do with the future, but retention and restitution were thought to depend on a number of integrated schemas. It was rather according to cultural frameworks, for example, that reconstruction was interpreted. However, what we learn from these data, is that the episodic memory system might be part of a more general system of mentally travelling in time, crucial for psychological continuity.

5.2. Shared-memory space – myself and others

Research on the fabrication of false memories nevertheless highlights the importance of the intersubjective contexts, or the importance of the relationship to others in the construction of memories. Elizabeth Loftus' research has shown that personal memory is malleable and that it is sensitive to a social environment. In the study that I have mentioned in Chapter 3, she induces subjects to remember a past event in an erroneous way. Let's recall the experiment here (Braun et al. 2002): after showing to two out of four groups of subjects a false Disney advertisement depicting Bugs Bunny, she asks them to tell about their memories of a day spent at Disneyland and asks them if they shook hands with Bugs Bunny. A large proportion

⁵⁸ On this point, see the discussion by Forest, 2014, and in particular Chapter 2, section 6, pp. 116 & sq.

of the subjects (30% and 40%) in the two groups that were in contact with the false advertisement state that they know or even remember shaking hands with Bugs Bunny during the day at Disneyland. But Bugs Bunny is not part of the Disneyland universe, so this memory is impossible; people have formed an induced false memory. The psychologist Julia Shaw has even shown that she could induce people to form the false memory of having committed a crime in their childhood ([Shaw and Porter 2015](#)). It was then impossible for them to distinguish between their false memories and their genuine memories. Loftus' hypothesis is the following: if one cannot distinguish between induced false memories and genuine memories, it is because memory, as a whole, is malleable. Interestingly, research on the fabrication of induced false memories emphasizes the relational character of individual memory. Individual memory is in relation to a world and to other individuals. We are not alone with our memories; they are also determined by collective memory. We can see that construction within this framework does not relate directly to our individual projections or longer-term goals, but rather to our relationships to others, to our environment, and to the social and interpersonal frameworks in which our memories are constructed and recombined.

In the case of induced false memory, this process might seem unbeneficial. But it could be the product of a beneficial system. It has been shown that remembering in groups could enhance memory performance in terms of the number of details retrieved and in terms of overall accuracy. The works on collaborative remembering ([Blumen, Rajaram, and Henkel 2013](#); [Harris et al. 2014](#); [Johansson, Andersson, and Rönnerberg 2005](#); [Sutton 2008](#)) have shown that joint remembering, especially in intimate groups, of a common past event, could enhance both the overall accuracy and the number of details retrieved, though this is not systematic ([Harris et al. 2019](#)). This seems to plead in favor of the idea that personal memories do enter a shared memory space, similar to what MacIntyre has described, where rememberers adjust their memories according to how the other members of the groups they belong to remember them. Whether this results in an enhanced accuracy or not depends on the quality of the information shared by all parties, and on the relations and the communication happening between them ([Harris et al. 2019](#)). I take this as evidence for the claim that episodic memory construction is sensitive to a social environment and to the social contexts of encoding and retrieval. In this regard,

MacIntyre's concept of shared narrative space could be adapted to understand this aspect of memory construction processes.

In psychology from the early 20th century to the present day, we have discovered that episodic memory is not only a preservative capacity, and that remembering the past in first person and imagining the future in first person are not such different abilities as we thought. But contemporary memory science, interested in the constructive dimension of episodic memory, has cut it off from its context, and it is surely this notion of context that needs to be rehabilitated in order to understand how construction in episodic memory works. It is now clear, therefore, that the relationship to the personal future is not the only determining ingredient in memory construction.

5.3. Constructive memory theory of personal identity.

The last question that I want to address is: ultimately, what does ground personal identity epistemically and metaphysically in a constructive memory theory?

5.3.1. Epistemic and metaphysical questions

How do I know myself? I have proposed in Chapter 1 that being a self, according to Locke's memory theory, is to have an epistemic constitutive relation to oneself. As episodic memory is both a preservative and a constructive capacity, it is able to fulfil this role. Remembering a past episode of my life is both relating to myself and constituting a representation of myself. Remembering is not only retrieving the past faithfully, but it is a flexible capacity of reconstruction of stored information. It corresponds to the construction of a representation of my past self, depending on stored traces of past experiences, on my present needs (related to action explanation and intentions) and on the contexts in which I experienced events in the past and I remember them now. In this sense, it can be constitutive from a metaphysical point of view: being a person is to recognize myself and to construct a representation of myself thanks to memory. It can also be constitutive in the detail: recognizing myself is to constitute the detail of my personal identity, what I have done, who I am, what I can do, what are my intentions.

5.3.2. Circularity again?

Psychological continuity is thus made by a mental action of self-recognition and self-extension to the past that does not presuppose it in the first place. Autonoetic consciousness does not presuppose personal identity because it can happen in cases where personal identity is not objectively true. One could object that I am confusing the sense of personal identity and metaphysical personal identity. My answer is Lockean. Being a person is to have self-consciousness and to recognize myself in different moments of time. It does not have to correspond perfectly to who I am as an individual. I am for myself who I believe I am. Though, as I have shown, my representation of myself is embedded in an environment that puts constraints on memory construction. And as my memories are formed in a shared memory space, I can be led to revise them, and to form representations of my past self that fit better to who I am as an individual.

5.3.3. What about self-deception?

But then, can false memories play the same role for my personal identity as genuine memories? I have put pressure on the notion of genuine memories and discussed the binarism between genuine and false memories. In chapter 3, I have argued that both the factivity condition and the causal condition (in a strong sense) are not necessary for remembering. The consequence of such a claim is that both genuine and false memories can play a constitutive role for my personal identity. They can both ground my psychological continuity, by constructing in the present a representation of myself extended to the past and as such constituting my personal identity.

5.3.4. What about extreme cases of self-deception such as in confabulation?

One could object that the account I propose leads to undermining the distinction between genuine remembering and clinical confabulation, and confabulation cannot make personal identity. Indeed, I do think that the distinction between

remembering and confabulating should be reduced in the context of a reflection on personal identity. Constructing a representation of oneself regardless of its accuracy allows the individual to self-extend beyond the present. A piece of evidence to support this claim is related to the literature on clinical confabulations. It has been shown that resistance to challenge is a common characteristic of clinical confabulations, and challenging has been characterized as non-beneficial in most cases (Fotopoulou 2008; Zangrilli et al. 2014), and in fact useless, since subjects rarely or never adjust their memories accordingly. The challenging of clinical confabulation is more likely to provoke a sort of distress in subjects and to enhance the distrust they have in other people. In the case where subjects cannot adjust their representation of the past to others' views, doubting it is harmful. When adjustment is difficult to make as it is the case in clinical confabulation cases, people nonetheless need recognition from others. They need their memories to be acknowledged.

In a sense, challenging clinical confabulations, despite the fact that the subjects cannot form another version of her past, might be seen as a denial of the subject's history, capacity to report it, and her authority on it. In a sense, in clinical confabulation cases, since people are unlikely to revise their memories, challenging them is to deny them the possibility of social recognition and acknowledgement of who they are. First, I take this as evidence for the claim that for a subject to maintain a psychological continuity, preservation is not necessary but the formation of a representation of oneself is needed. Second, I take this as evidence for the claim that another needed ingredient for psychological continuity is to be acknowledged by others, such that entering a shared memory space is important for psychological continuity.

5.3.5. Amnesic selves.

The last point that I want to discuss is the difficult case of amnesic selves for a constructive memory theory of personal identity. I believe that such a revised view of a memory theory changes some aspects of the problem. A memory view had to claim that amnesic selves cannot maintain a psychological continuity, though this

is misleading (Craver, 2012). Patient B (mentioned in 4.3) is a good example of this: he does not form a declarative memory of what happened in the hospital room he runs away from, though he does run away from it, and this can be seen as a consistent behavior given what happened to him in this hospital room. According to Craver, 2012, B can maintain a psychological continuity despite his episodic amnesia, and this should be a reason to abandon a memory theory of personal identity. As I have shown, it is not clear that B does not have a memory relation to this event. His behavior is in fact better explainable by memory, though he does not have a declarative memory of this event. This could lead us to rethink what amnesia exactly means, and whether people with episodic amnesia really do not have memory relations to their past at all. Another possible path is to argue that a constructive memory theory of personal identity describes one epistemic constitutive relationship that an individual can have to itself, but not all. Just as there are non-narrative selves, there could be non-memory selves.

Personal identity depends on a relation of self-recognition allowing self-constitution, and memory is just one of the means of such a relation. Some last remarks about amnesic selves. First, we need to distinguish cases of amnesia with confabulation and cases without confabulation. Confabulation can play a memory role and allow people to self-extend. Second, it is not exactly clear that people with episodic amnesia have no memory relations to themselves, even though such a relation does not display explicit memory. Third, since people are not isolated but embedded in a social context with people they trust, it is possible that those people can supplement their memories when they have difficulties in remembering the past.

Eventually, since I have shown that self-recognition (by oneself and others) is crucial for psychological well-being regardless of accuracy, it might be useful to change our practices in relation to people who confabulate, in the cases where they are not ready to revise their versions of the past. Instead of focusing on accuracy, and rejecting confabulations for what they cannot do, it might be beneficial to focus on what confabulations allow people to do, namely, construct a sense of psychological continuity.

Section 6 – Concluding remarks.

In this chapter, I have come back to Locke's theory of personal identity, and presented the missing ingredients of such a theory, grounding the reflection in particular on empirical memory science. I have wondered whether this should lead us to endorse a narrative view of personal identity and rejected this path. There are persons who do not or cannot narrate, and for this reason narrative capacities cannot be a necessary condition for personhood. The myth of life-coherence is misleading. And it is unsure that narratives are needed for a person to recognize herself and constitute her personal identity. However, I have proposed to adapt some of the concepts of MacIntyre's narrative view to a memory theory of personal identity. In particular I have contended that using the concept of shared memory space and memory adjustment could be helpful for a constructive theory of personal identity.

Eventually, I have sketched such a view. Personal identity corresponds to the construction of a representation of myself depending on past experiences, on my goals and intentions at the moment of retrieval and depending on the contexts in which I form such a representation. I have argued that such a view allows us to think that clinical confabulators could form such a representation and thereby, if acknowledged, have a sense of personal identity. Personal identity in this framework does not depend on truth or on preservation but on the construction of believed representations of myself. Constructive memory is one of the means of such a relation but does not have to be the only one.

The concept of constructive memory shakes up the memory theory of personal identity, since, if memory is at the foundation of the self, it is no longer in that it determines it definitively, but in that it rewrites it, and reconstructs it. In this regard, the concept of constructive memory invites us to overcome the antinomy between memory and narrative theories, since the autobiographical memory thus understood can be thought of as including a social background and as involving a form of construction. Moreover, it seems that the concept of constructive memory can clarify the notion of self-narrative and shed light on how it can be a principle of construction of unity and continuity of personal identity.

I have therefore made an attempt to propose a concept of the self that allows us to better think about our effective reality, that is to say, that understands it in a personal time, anchored in a social time and in a world, and that allows us to think about the continuity of a person with herself, even in the borderline cases of confabulation or amnesia.

Conclusion

In this thesis, I have proposed to rethink memory in order to rethink personal identity. I started from an interrogation about personal identity. How is it possible that people, despite the changes that affect them, recognize themselves as themselves? How is it possible that they generally do not doubt who they are, in the sense that they generally do not doubt being continuous with themselves? How is it possible that we live our lives in a pretty fluid way and that most of the time the continuity of the self, the psychological continuity, is not even an issue?

A common answer to the question of personal identity is that memory is what makes the self: memories allow us to connect with ourselves and to have an idea of ourselves that extends beyond our present stream of consciousness. It is thus episodic memory that produces psychological continuity.

But what also interested me, particularly at the beginning of this thesis, was the recent research on the constructive dimension of episodic memory and its possible relation to the imagination of the future in a first-person perspective. In 2007, researchers in neuropsychology formulated the constructive episodic simulation hypothesis (Schacter and Addis, 2007). This hypothesis, which I encountered early on, motivated this research.

If episodic memory is constructive, although it does not seem to be, and if it is essential to our representation of ourselves and our psychological continuity, what is it to recognize ourselves as the same person through time? In short, what difference does it make, to the question of personal identity, that memory is not what we thought it was?

I have been interested first in a tradition that considers memory as the essential ingredient of personal identity. This tradition begins with John Locke. In the first part of the thesis, I explored the Lockean theory of personal identity. And from the outset, I supplemented my reflection with contemporary empirical research on memory. Locke's idea is the following: what is essential to recognize oneself as a self from a synchronic point of view is consciousness, in the present, in order to individuate oneself; and from a diachronic point of view, the consciousness of being the same, still in the present, in order to recognize oneself as the same at different moments of time and to give oneself a temporal extension. In this second dimension, I discovered and defended, although it is the matter of a debate, the importance of a memory capacity in Lockean theory. But with it came new problems, some that have been raised for a long time, such as the circularity and transitivity of memory, and others more recent, related to the constructive dimension of episodic memory discovered by contemporary science. The first part of the thesis is thus a contribution to Locke's scholarship. In particular, it is an attempt to explore Locke's theory in light of what we know today about memory.

Memory and personal identity in Locke

Personal identity and circularity

In the first chapter of the thesis, I have sought to describe Lockean theory, relying in particular on chapter 27, and discussed its interpretations. About personal identity, two types of questions are traditionally asked. An epistemic question: what

makes it possible to recognize oneself as oneself? And a metaphysical question: what makes the self? What is the basis of personal identity?

These two questions have their source in John Locke's philosophy. And among Locke scholars, there is debate as to which of these two questions he sought to answer. Weinberg, for example, defends a metaphysical interpretation of the Lockean theory of personal identity. According to her, there is in Locke a metaphysical fact of consciousness, which is inaccessible to us and yet is the basis of personal identity. Consciousness has two dimensions. A subjective one, which allows me to identify with myself, and an objective one: the same continued consciousness, as an objective fact, grounds personal identity. Newman, on the contrary, proposed an epistemic reading of Locke's theory, according to which Locke defended a thesis about self-recognition and self-access, without pronouncing on metaphysical questions.

I tried to overcome the dichotomy between epistemic theory and metaphysical theory of the self. It seemed to me that a faithful reading of Locke, on the contrary, invited us to consider that there is no separate answer to one of these two questions, and that the interest of Locke's intuition was to answer the two questions, taken together. And more precisely to answer the metaphysical question of personal identity with the question of self-recognition. To be a person, in Locke's view, is to recognize oneself as the same at different moments of time and thus, in this act of self-recognition, to constitute oneself as a person with a temporal extension.

But Locke's theory of personal identity has been the object of criticisms, including an objection made famous by Butler: self-consciousness, at the heart of self-recognition, presupposes, and therefore cannot constitute, personal identity. This objection has received many responses in the literature. In this thesis, I wondered whether the reading I was proposing was threatened by the objection of circularity. I wondered whether the recognition of personal identity could happen in the absence of a personal identity preceding such recognition. The alternative I proposed was that either the recognition of personal identity is always only an act of recognition of something that is already, and in this case, recognizing oneself presupposes personal identity, or the recognition of personal identity can occur

without a personal identity preceding the act of recognition, and in this case, the recognition of oneself can ground and constitute personal identity. Then, self-recognition does not presuppose the identity of the person.

I proposed to consider the case that Locke calls "the fatal error": an individual wrongly remembers having done an action that she never performed. But by remembering it as her own, she identifies with the subject of that action, recognizes herself in it, and considers the consciousness of the remembered to be the same consciousness as the rememberer's.

The passage in which Locke considers the fatal error is the matter of a lively debate in Lockean studies. For many commentators, this passage is evidence for the claim that Locke's theory is inconsistent. The fact that I may be mistaken about my personal identity implies that it definitely cannot depend on my access to it.

However, I have argued that, in Locke's view, being wrong about oneself as a person is an ambiguous issue. If we give full weight to the distinction he makes between individual identity, human identity and personal identity, it is no longer an issue. In the case of the fatal error, I can be wrong about my self-identification as a human individual, but I cannot be wrong about my personal identity, since this depends on my access to it in a first-person perspective. In the case of the fatal error, in fact, I do not make a mistake in terms of my personal identity. By identifying myself with a person in my memory, I constitute my personal identity with that of the remembered person.

Locke invokes God in this passage and, as if anticipating criticism, he says that at Judgment Day, God will resolve these errors of identification by practicing what I have described as a kind of Platonic anamnesis: God will make each person aware of his or her actions, and thus individual identity and personal identity will coincide. What is interesting is that even when Locke invokes God about personal identity, he does not attribute to him the role of deciding objectively on people's identities, but only of restoring to each person the consciousness of what she has done. In other words, even in this case, the answer to the question of personal identity must pass through the consciousness of the subject, and the answer to the metaphysical

question depends on the answer to the epistemic question. Personal identity depends on its recognition by the subject.

I proposed an empirical argument against the objection of circularity and brought the case of the fatal error closer to what is called today in psychology false memories. I have shown that these cases are not science fiction, but common cases where a subject remembers having done an action that he or she did not perform. However, in these cases, the subject who remembers identifies with the subject he or she remembers. I introduced the concept of auto-noetic consciousness, proposed by Tulving, 1985 to understand these cases: auto-noetic consciousness is the type of consciousness that accompanies mental time travel in first person. Memories are accompanied by this type of consciousness, according to which the subject who remembers is also the subject of his or her recollection. In cases of false memories, regardless of the accuracy of the memory, the memory is also accompanied by an auto-noetic consciousness, which is even why I have the impression that I remember. But these are cases in which there is no personal identity preceding the act of recognition. Since cases of fatal error, or false memories, are possible, and since in these cases, even if I wrongly identify myself with the subject I remember, I still identify myself, then recognizing myself as the same and extending myself through time does not presuppose personal identity, but can produce it. I have therefore shown that auto-noetic consciousness, that is, the consciousness of personal identity over time, since it accompanies memories and false memories, does not presuppose personal identity.

Thus, I contended that consciousness and memory were essential to Locke's theory of personal identity, as capacities that allow me to recognize myself and thus to constitute myself. Therefore, the first chapter offers a reading of Locke that goes beyond the traditional dichotomy between the epistemic and the metaphysical question of personal identity.

Memory and the preservative view

In the second chapter of the thesis, I continued this exploration of Locke's theory of personal identity by questioning its conception of memory and the role that memory can play in personal identity. This chapter thus opens with a study of the passages in which Locke outlines his conception of memory and the capacity to remember (in particular Chapter II, x of the Essay). I have shown that Locke has a preservative conception of memory. In Locke, memory is supposed to preserve the ideas one has had, in order to be able to retrieve them later. I tried to reconstruct a theory of memory in Locke's text and listed the conditions of memory. For a mental state to belong to the category of memory, it must meet several conditions:

- (1) Representation: An individual A now has a representation of a past event E in t_1 .
- (2) Previous awareness condition (PAC): A must have had the idea of E in the past at $t-1$.
- (3) Self-awareness: A is aware that his representation of E in t_1 is the representation of an event or idea he previously perceived in $t-1$.
- (4) Sameness: The representation of E in t_1 by A is the representation of the past and preserved perception of E in $t-1$ by A. The representation of E now and in the past tense must be the same idea.
- (5) Accuracy: The representation of E in t_1 is a true representation of what A perceived in $t-1$.
- (6) Storage: E must have been stored when E was perceived, in the form of a memory trace.
- (7) Causality: The representation of E by A at t_1 is caused by the perception of E by A at $t-1$, thanks to the preservation of its impression.

The storage condition (6) and the causal condition (7) are at the heart of the Lockean conception of memory, since they secure together the sameness condition and the

accuracy condition. Together, these conditions ensure that memory is the recollection of faithful representations of the past and can therefore constitute personal identity in Lockean theory.

But a second famous objection, by Berkeley and Reid, puts pressure on his theory: grounding personal identity on memory leads to a theoretical problem of transitivity. Reid's version, known as the "brave officer's" version, is as follows: suppose a brave officer who was beaten as a child in school for stealing fruit, who, as a soldier, took a standard from the enemy in his first campaign and was later made a general. Suppose also that when he took the standard, he remembered stealing the fruit. And suppose that when he was made general, he remembered taking the standard but forgot he stole the fruit. According to Locke, he is the same person as far as his self-consciousness extends. It follows that when he was a soldier, he was the same person as the child thief, and when he was made a general, he was the same person as the soldier. However, when he was made general, he cannot be described as the same person as the child thief since he does not remember this past experience. Thus, he is and is not the same person as the child thief. Locke's theory therefore suffers from a transitivity problem.

I have been particularly interested in Reid's formulation of the transitivity objection, because it is accompanied in Reid's work by a more general critique of Locke's conception of memory. This is not only a critique of his theory of personal identity but also a critique of his conservative conception of memory. In Locke's view, when I remember, I relive a past experience by retrieving an idea that was previously formed during perception and stored. For Reid, on the contrary (and this is part of his metaphysical background of defending direct realism), memory cannot be the recollection of past experiences but is a direct relationship to events experienced in the past. He argues against Locke's conception of memory as a storehouse. First, the existence of impressions stored in the brain lacks evidence. Moreover, even if such traces were possible, they could be mere correlates and not causes of memory. Finally, since Locke says that ideas cease to be when they are not actually perceived, the sameness between perception and memory ideas is unlikely. According to Reid, a thing cannot have two beginnings of existence, and furthermore, he accuses Locke of not distinguishing between perception and memory.

In order to address these two criticisms, I first studied the paths followed by Locke's scholars to overcome the objection of the brave officer. They are of two types. Philosophers have tried to adapt the concept of memory by arguing that memory binds ideas by a relation of ancestrality. The general, by identifying himself with the soldier, is linked to the experiences that the soldier has appropriated even if he no longer remembers them. A second strategy consists in discounting the importance of the role of memory in Lockean theory. This path has been proposed by Atherton (1983), for whom what makes personal identity in Locke is consciousness, not memory. It is followed by Shelley Weinberg (2012), and also in a sense by Hamou (2014).

Their reading has the merit of solving many of the problems of Locke's theory. But I have tried to show that they face difficulties in explaining passages where Locke seems to consider memory and extended consciousness as one and the same thing. And without memory, it is not clear how a subject can access the details of his personal history. Yet, as I have argued, self-access is at the basis of Locke's conception of personal identity.

Finally, the second part of Reid's critique is little considered in the literature, and I have shown that the two paths taken by Locke scholars share a conservative conception of memory with Locke.

In the final sections of Chapter 2 I have proposed a third path to address Reid's objection, based on recent empirical findings. I proposed to rethink the role that memory may have in a Lockean theory. Against Reid, and drawing on memory science, I argued that there are such things as memory traces. But they do not necessarily imply the truthfulness of memory, and even less do they guarantee the sameness condition. I thus proposed a hybrid view of episodic memory, as both a preservative and a constructive capacity.

I first showed that the psychology of memory since the end of the 19th century provides elements in favor of the Reidian critique of the preservative conception of memory. Memory does not give us access to accurate images of the past. I have

drawn on the research on eyewitness testimony that began in the late 19th century and continues to the present day (Loftus, 2005), which shows that eyewitnesses never remember a scene perfectly. They add elements, forget others, and reconstruct a representation of the event. Attending an event does not imply that one remembers it truthfully.

I have argued that both encoding and retrieval are constructive processes. Perceiving a scene is not perceiving the truth of a scene, it is always a look at a situation that depends on the subject's background, bias and attention. To show this, I have drawn on the experience of Sophie Calle, on empirical research (Mulligan & Hornstein, 2003), and on the philosophical literature on the constructive dimension of encoding (McCarroll, 2018).

I also showed that memory is not only constructive, but also reconstructive. That is, each recollection is a process of reconstruction that gives a new representation of an event. Bartlett's studies and his experimental method of Repeated Reproduction provide an illustration of this (Bartlett, 1932). I argued that Reid was right to put pressure on the sameness condition, but that this did not imply adopting a direct realism. Rather, it required adopting a constructive conception of episodic memory.

If I showed that episodic encoding and retrieval have a constructive dimension, I also defended that this was not all. And I questioned Reid's conclusions about the nature of memory. Memory does not put people in a direct relation with past events but functions as the retrieval of stored and encoded mnemonic traces. I relied on the literature in psychopathology that has been published from the early 20th century to the present day in the study of amnesic patients. A famous case is that of HM who, after a lobectomy, sank into a deep amnesia, especially with regard to his recent memories. He also had difficulty forming new memories. The case of HM is not isolated. In episodic amnesia, a part of the brain called the hippocampus is often damaged and damage to the hippocampus always results in a form of episodic amnesia. The hippocampus is necessary for episodic memory. While in Reid's time there was not enough evidence to support that memory has to do with the brain, the situation is different today.

I have given reasons for adopting a hybrid view of episodic memory that understands it as both preservative and constructive. I have rejected a purely preservative view of episodic memory and rejected the condition of sameness in a memory theory. I therefore hypothesized that Locke's preservative view of memory was unlikely, but that a hybrid view of memory, on the other hand, provided a better understanding of the fundamental role of episodic memory in personal identity, and avoided the classic objections. The specificity of the reading I proposed is thus to reconsider Lockean problems in light of contemporary memory science.

In the second part of the thesis, I deepened my reflection on the concept of memory. This exploration is part of a contemporary reflection in the sciences and philosophy of memory on memory capacity, and in particular on episodic memory. In chapter 3, I questioned the necessity of two conditions present in most contemporary memory theories: factivity and the causal condition. In Chapter 4, I looked at the functions of episodic memory and explored its non-epistemic functions.

The constructive dimension of episodic memory

The conditions for remembering

Chapter 3 explores contemporary theories of memory. In particular, I have discussed the factivity condition and the causal theory of memory proposed by Martin and Deutscher (1966) and adopted in most theories of memory.

According to the factivity condition, a subject *S* remembers an event *E* if *E*. For a subject *S* to have a memory *M* of an event *E*, *E* must have occurred in the past. This condition ensures that the memory is the memory of a true event.

According to the causal condition, for subject S to have a memory M of a past experience Exp, Exp must have been operative in the production of M. The causal condition guarantees that the memory is an accurate memory of a past experience.

In this chapter, I have proposed a fictional case, which on the one hand is part of a tradition in philosophy of mind of which Locke is again one of the initiators, and of which neo-Lockeans, such as Parfit for example, are fond. But, on the one hand, this case is more like a police scenario (a murder in a retirement home, for which Anna, a nurse, wrongly believes she is responsible) and, on the other hand, I did not use it to draw theoretical consequences from intuitions. I used it to isolate different cases in which the question of whether mental contents are memories arises. I therefore proposed to consider 4 cases:

M1: Anna remembers having given Ms. F. sleeping pills. This event did occur, and her experience may have produced her memory. M1 can meet both conditions.

M2: Anna remembers giving hydroxyzine to Ms. F. However, Anna did give Ms. F sleeping pills but not hydroxyzine (the factivity condition is not met). Anna's current representation is not only caused by past experience but also by information she received afterwards: she found an empty hydroxyzine tablet on Ms. F's bedside table the next day (the causal condition is not strictly fulfilled).

M3: Anna remembers seeing Ms. J in her doorframe the next morning when she got rid of the empty hydroxyzine package in the hallway. Anna actually hallucinated; Ms. J was not in her doorframe; Anna saw a shadow and thought it was Ms. J (the factivity requirement is not met). Anna's current representation may be caused by her past experience, though (the causal condition is met).

M4: Anna remembers giving anti-allergy pills to Ms. F. That is what happened (the factivity condition is met). Her current representation is the result of a memory revision. It is caused by post-event information and perhaps also by the recollection of the past experience (the causal condition is not strictly met).

If the two conditions discussed are necessary for remembering, then Anna genuinely remembers in M1 only. In other cases, it is something else. The challenge of Chapter 3 was to discuss the necessity of these two conditions.

I started with a discussion on factivity. I argued several things. On the one hand, it implies a form of disjunctivism which, in addition to lacking evidence, is difficult to understand in the case of M3. M3 is a case in which Anna remembers, perhaps wrongly, but she does remember. The factivity condition makes no room for memory errors and this is a problem.

On the other hand, in memory theories that posit factivity as a necessary condition, it is often correlated (as in Cheng and Werning, 2016) with the idea that false memories are not memories. They are not memories because they do not respect the factivity condition. I have shown that this idea is based on a misunderstanding. Research on false memories, in fact, does not allow us to decide whether false memories are true or false in relation to the past event they represent. Rather, they are related to the causal condition, since false memories are the result of the integration of post-event information into the representation of a past experience. The information in question can be misleading (as in M2), but it can also be true (as in M4) and allow for a revision of a distorted memory. While the conditions of factivity and causality are often presented as interdependent, I have shown that they are not necessarily fulfilled together. The problem of false memories is more related to the causal condition than to the factivity condition. Memories resulting from the reconstruction of past events with post-event information integration are not always false.

Finally, I have argued that positing factivity as a necessary condition for remembering actually presupposed a biased conception of episodic memory. The factivity condition is not necessary for remembering, it is necessary for knowing the past on the basis of memory. In other words, considering the factivity condition as a necessary condition for remembering implies that one has already decided what episodic memory is for: the knowledge of the past. I contended that this presupposition should be questioned (and this is the subject of Chapter 4).

In the third section of Chapter 3, I discussed the causal condition. The causal theory of remembering is endorsed by most contemporary theories of recollection, but it has recently been the subject of debate. It was first proposed by Martin and Deutscher (1966). And I have explained it as follows:

Causal condition: The memory M of S of the past Exp experience must be appropriately caused i.e. by a non-deviant causal chain, by Exp.

According to the causal condition, M2 and M4 are not memories. Whether true or false, they are not appropriately caused by the past experience, but have another source. M1 can be defined as a case of memory, because it can be caused by past experience. So can M3. Although it is false with respect to the past event, it can be authentic with respect to the past experience. Anna hallucinated, but she did have that experience. Her hallucinated memory could be caused by her past experience.

I have shown that in M4, according to the causal theory, Anna believes that she remembers but in fact makes a "source-monitoring error" (Johnson et al. 1993). That is, she forms an erroneous metacognitive judgment about the source of her representation of the past: she believes that it comes from her past experience, when in fact it comes from post-event information. M2 is the result of a misinformation effect. Its source is not the past experience, but it is caused by post-event information. M4 is true (with respect to the past event) but inauthentic (with respect to the past experience); M2 is false (with respect to the past event) and inauthentic (with respect to the past experience).

In the section on the causal condition, I argued that the binarity between false and genuine memory does not help to explain episodic memory.

I first showed that it is common to remember more than what was present in a represented past experience, as shown in particular by the DRM paradigm (Roediger and McDermott, 1995). The normal function of episodic memory involves an activity of gap-filling, of recomposition. The idea that the past experience must be the cause of memory is not really compatible with this understanding of episodic memory.

I have shown that both encoding, and retrieval are processes that rely on the gathering of information from various sources. On the one hand, encoding is not only caused by past experience, since perception involves completion processes, as is the case in the robust phenomenon of boundary extension. On the other hand, episodic memory retrieval also relies on the recollection of information from various sources. In particular, it also relies on semantic information, on what the remembering subject knows.

I have argued that episodic memory interacts with other types of information, so it is difficult to say that a recovered episodic memory comes only from the past experience it represents. Either one endorses a weak version of the causal condition, that the memory of Exp must be partially caused by Exp, but in this case it is not very explanatory. Indeed, according to this weak version, M1, M2, M3, M4 can count as memories, they can all be described as partially caused by the past experience they represent. In a sense, such a condition would lose its explanatory power. Or, one endorses a strict version of the causal condition, and in this case, in fact, none of M1, M2, M3, M4 can really count as a memory. They can all be described as resulting from the recollection of information from various sources. Such a condition would also lose its explanatory power.

I therefore argued that the factivity condition and the causal condition (in its strict version) could not be considered as necessary conditions for remembering. The third chapter is thus a contribution to the current controversies in philosophy of memory about factivity and the causal theory.

Non-epistemic functions of episodic memory

These two conditions play an important role in epistemic theories of episodic memory because they are supposed to guarantee, for one, the truth of the memory in relation to the past event it represents (factivity), and for another, the authenticity of the memory in relation to the past experience it represents (causal condition).

But they presuppose that the main function of episodic memory is to represent the past in a true and authentic way.

In Chapter 4, I discussed this presupposition. The question of the functions of episodic memory is also the subject of a lively debate in recent research on memory in cognitive science and philosophy. Researchers have recently argued that the discovery of a constructive dimension of memory, by focusing on memory errors, overlooks what episodic memory is used for: knowing the past. According to them, a central question about memory is its evaluation, and memories are evaluated according to their epistemic value on the past.

I began the discussion by explaining their arguments. I showed that this thesis has been defended in three ways in recent literature. In Mahr and Csibra (2018), the focus is on the causal condition. According to them, research on the constructive dimension of episodic memory is linked to a descriptive project within psychology, which describes what memory does. But this should not compete with what memory is supposed to do. There are two reasons for this: on the one hand, memories present themselves to the subject as being caused by their past experience. The memory (output) appears to be caused by a past experience (input). On the other hand, there are two different ways of asking the question about the nature of memory. One is descriptive, focusing on the best description of episodic memory. The other is normative and focuses on the conditions that justify considering a mental state as a memory. If one understands that the causal theory is engaged in a normative and not descriptive project, then the empirical arguments against the causal theory do not hold.

A second way of supporting this thesis is proposed by Henry and Craver (2018) and Craver (2020): normative memory has conditions of success that do not depend on empirical memory. Normative memory has conditions of success that are epistemic. And the number one rule is that a memory corresponds to first-hand experience. A rememberer has a specific epistemic authority: she was there, and she remembers it. While empirical research shows that memories can have sources other than first-hand experience, this does not change the fact that memory evaluations depend on them. First-hand experience has an important epistemic value.

A third path has been proposed by Barner (draft) and focuses on normative discontinuity. Although empirical research on mental time travel leads some to adopt continuism (Michaelian, 2016a) and to consider past oriented mental time travel and future oriented mental time travel as being of the same kind, there is a fundamental normative discontinuity between remembering the past and imagining the future. Memories and simulations of the future cannot be evaluated epistemically in the same way. Memories are factive, they have occurred, they have truth conditions. Simulations of the future, on the other hand, have not happened yet. Memories are evaluated according to their truth with respect to the past, but simulations of the future cannot be evaluated in the same way.

Against these three versions, I have claimed that episodic memory does not have only epistemic functions. These arguments about normative memory and empirical memory hide the fact that episodic memory is not entirely oriented towards the knowledge of the true past.

My first argument against what I have called the epistemic-normative thesis is that the standards for memory evaluation are not unique. I have explored cases of clinical memory distortions. I have argued that in cases where memory distortions are epistemically innocent (Bortolotti, 2015), that is, where the person remembering or confabulating could not form a more authentic representation of the past, they may have epistemic benefits. In these cases, the standards of their evaluation are not the same as in so-called normal cases. Other rules apply that are not those of factivity or causality. I have drawn on the literature which shows that clinical confabulations have benefits in at least two directions: they prevent the harmful absence of self-representation and confidence in one's abilities, and they are a way of training epistemic skills (constructing a representation of the past, formulating it, confronting it with the judgment of others in order to possibly revise it).

My second argument was to show more generally that memory science, by discovering its constructive dimension, has invited us to revise the ways of evaluating it epistemically. I contended that, since memory construction depends on the contexts of encoding and retrieval, personal memory is not sufficient for the

knowledge of the past. I argued that the empirical domain can guide the normative domain. If we want to know the past on the basis of memory reports, this must take the form of an inquiry.

My last claim was to argue that the epistemic regime is not the only normative regime that guides memory and memory evaluations. I have shown that episodic memory is not evaluated in all contexts according to its capacity to represent the past faithfully. There are, on the contrary, contexts in which the epistemic function of episodic memory is not what matters. If this is the case in a legal context, for example, it is not the case all the time. In the last section of Chapter 4, I explored non-epistemic episodic memory evaluations: memory evaluations related to psychological benefits and potentialities of action. I argued that episodic memory has non-epistemic functions.

I relied again on research on confabulations, and I showed that memories, regardless of their (in)accuracy, that is to say, also regardless of their epistemic cost, could have psychological benefits. Attempts to challenge clinical confabulations are almost always unsuccessful and not beneficial (Fotopoulou, 2008). Despite their epistemic cost, researchers have shown that clinical confabulations have psychological benefits. In particular, confabulation can improve self-confidence, a sense of competence, coherence and psychological well-being (Bortolotti and Sullivan Bisset, 2018).

I have also proposed that certain memories should not be evaluated according to their fidelity to the past but according to their effect in relation to potentialities of action. I have proposed two childhood memories to illustrate this argument.

One is a memory of failure, the other is a memory of success. I have shown that one is at the basis of my sense of competence in certain tasks, and the other is at the basis of my sense of danger in certain situations. What matters to me in these memories is not their accuracy but what they allow me to do, and how they form the basis of my personal identity and sense of competence.

In this sense, the arguments that I have proposed in this chapter are not only directed against the epistemic-normative thesis, but also more generally against the epistemic theories of episodic memory that are in fact dominant in the emerging contemporary field of philosophy of memory. Even in Michaelian (2016c), who offers one of the harshest critiques of the causal theory of memory, episodic memory is oriented towards the reliable knowledge of the past. I have tried to show that what these positions have in common is to presuppose that the main function of episodic memory is the knowledge of the past. I have argued, on the contrary, that this is not its only function. The fourth chapter is therefore a contribution to the reflection on the functions of episodic memory. If contemporary philosophers mostly consider episodic memory for its epistemic functions, I have shown that episodic memory also has non-epistemic functions.

Constructive memory and personal identity

In the last part of the thesis, I returned to the question of personal identity. I tried to update what a memory theory of personal identity could be if we understand the constructive dimension of episodic memory. Indeed, the image of episodic memory that emerges from empirical research is the image of a constructive capacity, dependent in particular on the contexts in which it is constructed. I proposed to return to the Lockean theory explored in the first part of the thesis and take stock of the missing pieces of his memory theory in the light of what episodic memory in fact does.

I first showed that Locke did not take into account the constructive dimension of episodic memory. Locke has a preservative view of memory, and this is a problem for both his theory of memory and his theory of personal identity.

I then argued that Locke's theory reduced the self to its individuality. It posits a person independent of a world and a context, an isolated person alone with his or her consciousness. This criticism, already formulated by Leibniz (*Nouveaux Essais*, II; xxvii), resonates with what we know today about personal memory. Memories

are formed and retrieved in a social world. Not only are they sensitive to the contexts in which they are formed and retrieved, but in addition, people who remember are rarely alone when they remember, and were rarely alone in their experiences. In these cases, the same event may be remembered by more than one person. And when we collectively remember the past, we commonly adjust our memories to the versions of others. There is a social dimension of remembering that is missing in Locke's theory and yet essential to memory processes and personal identity.

Finally, I argued that the issue of personal identity is not only related to self-extension to the past, but also to my survival in the future. Drawing on the literature in psychopathology that explores the consequences of episodic amnesia, I showed that psychological continuity depends not only on remembering the past in first person, but also on projecting oneself into the future in a first-person perspective. A problem for patients with episodic amnesia is also that they have difficulty projecting themselves into the future, as Tulving (1985) has shown. This dimension of apprehension of the future, the importance of the relation to oneself in an extended subjective time from the past to the future of the person, is also a missing piece in Locke's theory of personal identity.

In section 2 I explored the narrative theory of personal identity proposed by MacIntyre (1981) for the following reasons: on the one hand, the missing pieces of Locke's theory are present in MacIntyre. On the other hand, if personal identity depends on memory and memory is a constructive capacity, dependent on projections into the future, and with an important social dimension, it may be closer than expected to a narrative capacity. I have focused on MacIntyre's narrative theory more than on more recent versions of narrative theory. As such, it is possible that some of the objections I make in the chapter may not be valid for all narrative theories. But I have focused on MacIntyre for at least two reasons: first, because he is one of the first to propose such a view of personal identity. Second, because some of the concepts he formulated, such as the shared narrative space, narrative embedment, and narrative adjustment, were of particular interest to me. I consider that they can be useful for a better understanding of construction in episodic memory, as well as the constitution and recognition of personal identity.

On MacIntyre's view, people's identities are based on the fact that episodes in their lives belong to a larger whole that they can relate to and in which they are embedded. The unity of a life depends on the intelligibility of a person's actions, in their coherence within a shared narrative space. But narratives are never fixed. There is always a multitude of possible descriptions of a sequence of actions, and different versions can be given at different moments in time depending on what needs to be explained more specifically. A narrative capacity is therefore a constructive and reconstructive capacity.

There are thus a multitude of possible self-narratives, and the best narrative is constructed in a dialogue. Moreover, it is always embedded in the frames in which it takes shape. I have thus shown that the narrative theory responds to the individualistic inadequacy of Locke's theory. My self-narrative is embedded in the narratives of my social context and my time. These frames themselves have their own history. How I construct my self-narrative depends on the contexts in which it is formed.

Beyond the contexts in which my self-narrative is embedded, I am also included in a social world. Others can add things to my narrative, as Leibniz (*New Essays*, II, xxvii) has suggested; they can also produce a narrative of myself that is different from mine, and thus lead me to modify it. When I produce a self-narrative, it enters a shared narrative space in which people can give and ask for reasons for what they have done. A self-narrative is always subject to revision and adjustment. The construction of the self in a self-narrative is therefore not an individual activity but is embedded in a social world.

I have shown that a final essential ingredient in MacIntyre's narrative theory is the concept of intention. According to MacIntyre, for episodes to be episodes in my life and my history, they must make sense in the continuity of my intentions. Those intentions may change, but for past episodes to be intelligible to me and others and to form the basis of my personal identity, they must be explainable in a self-narrative and in the context in which they took place. Along with the concept of

intention, MacIntyre's theory provides a broader understanding of psychological continuity than just self-extension to the past.

While MacIntyre's theory seems to compensate for the shortcomings of Lockean theory, I have argued that they have common assumptions. The concept of narrative appears as a radical alternative to that of memory, yet it seemed to me that it was this alternative that needed to be overcome. That which opposes the truth of memory to the constructivism of narrative, continues to presuppose that memory is a capacity of true recording of the past. And yet, as I have shown in previous chapters, the image of episodic memory that emerges from contemporary research is that of a constructive capacity.

In section 4 of the chapter, I gave reasons for not adopting a narrative theory of personal identity. I recalled Strawson's arguments and argued that the fact that some human individuals do not have a narrative relationship to themselves is sufficient to refute a narrative theory of personal identity.

I proposed two additional cases from the empirical literature, however, which I put forward as arguments against the narrative theory: the case of nonbelieved memories and the case of B, an amnesic patient described by Störing.

I first attacked the idea of the coherence of the mental life by relying on the literature on belief. I showed that people have contradictory beliefs and may even have nonbelieved memories, i.e., episodic memories that are not accompanied by the belief that what I remember happened to me. The case of nonbelieved memories is difficult to understand, since episodic memories seem to be necessarily accompanied in their phenomenology by the belief that they are true. I have shown, however, that if we abandon the myth of the coherence of mental life, we have a better chance of making sense of such a phenomenon.

The second case I have proposed is that of B, a patient of Störing, discussed by Craver (2012). Craver presents B's case as a case of a person with episodic amnesia. He uses this case against memory theories of personal identity and to show that individuals with episodic amnesia can maintain a sense of self and a psychological

continuity. But the case he describes seems to me to be a case of episodic memory. It seems to me that it is a case in which B has an episodic memory with a behavioral, though not declarative expression. The case is as follows: B enters a hospital ward where he has had a traumatic experience in the past. And although he has no declarative memory of that experience, he runs away. I have shown that in this case, B actually has a behavioral response that is related to the memory of a specific past episode, as in episodic memory. It is not a conditioned response that he has learned, it is related to a particular event. I argued that there might be cases of non-declarative episodic memory. In this sense, I showed that an individual could maintain a form of continuity with herself without explicit memory and without narrativity, but I contended that the memory relation to oneself was important.

In the last section of the chapter, I tried to propose a theory of personal identity that would be based on constructive memory. I revisited the epistemic and metaphysical questions of personal identity. I showed that if memory is constructive it can play the constitutive role that Locke attributes to it by being a flexible capacity for reconstruction and recomposition, according to preserved traces of past experiences and their recombination according to the contexts of the present situation (what I am trying to explain, what I am trying to do). To recognize oneself through memory is thus to constitute oneself. I have thus shown that a concept of constructive memory was in fact more likely to play a grounding role in a Lockean theory of personal identity. Memory is constitutive from a metaphysical point of view: to be a person is to recognize oneself and to construct a representation of oneself thanks to memory. It is also constitutive in the detail: to recognize oneself through memory is to constitute oneself in the detail of one's personal identity, what one has done, who one is, what one can do, and what are one's intentions.

I have argued that such a theory no longer falls under the objection of circularity. I showed that false memories could also play a constitutive role in personal identity. I argued that even clinical confabulations could play this role. I claimed that, in the context of a reflection on personal identity, there was good reason to reduce the distinction between remembering and confabulating. I relied again on the literature on confabulation. Questioning clinical confabulations is often harmful. Resistance to doubt is even often considered a characteristic of clinical confabulation. But

questioning them, in addition to being useless since subjects rarely revise their representations of the past, is not beneficial. It has been shown that it more often causes distress in subjects and worsens the distrust they have in others. In cases where subjects cannot adjust their representations to those of others, questioning them is harmful. This seems to show that confabulations play a role in self-representation and self-constitution, and that recognizing them and having them be recognized is essential to psychological continuity.

I finally discussed the question of amnesiac selves. I argued that a revised version of a memory theory, which takes into account the constructive dimension of memory, changes something about the problem. Memory theories had to argue that people with amnesia cannot maintain a psychological continuity, which is a difficult thesis to defend. Craver (2012), on the contrary, showed that people can maintain psychological continuity despite episodic amnesia.

I argued several things in response. On the one hand, one must distinguish between cases of amnesia with and without confabulation. Confabulation can play a memory role and allow people to expand, and it is not clear that in cases of episodic amnesia, people do not have a capacity for self-recognition at all based on stored and recombined information. Secondly, since people are not isolated, but understood in a social context populated by people they trust, it is possible that they can make up for the missing memories. Finally, since I have shown that self-recognition is crucial regardless of accuracy for psychological well-being, I have claimed that it might be useful to change our practices towards people who confabulate. Rather than focusing on fidelity to the past and rejecting confabulations for what they cannot do, it might be useful to focus on what confabulations allow people to do, i.e. to build a sense of psychological continuity.

I thus claimed in this thesis that personal identity does not depend on the truth of oneself, but that it relies on a relation of recognition of oneself, of one's past and of one's intentions, which constitutes persons. A constructive memory can play this double role: epistemic and metaphysical. It is at the same time a capacity which allows me to recognize myself and, because this recognition is not a simple recognition of an object which precedes it, it can constitute me, as a person with a

history, capacities, intentions. So, I have in fact shown several things: on the one hand, that the concept of memory had to be rethought in a memory theory of personal identity. On the other hand, that the concept of person also needed to be rethought. The person is not an object whose content would be fixed independently of its apprehension by a subject. This apprehension produces something, it produces its object. Personal identity is the result of a constitutive recognition of oneself. I contended that the question of the truth of memory as well as the truth of the self had to be overcome. They can in fact be secondary in a reflection on personal identity. While questions of truth, accuracy with regards to the past, and successful remembering are central to discussions of memory, and are presented in the literature as essential characteristics of episodic memory, I have shown that they are not essential in all contexts of reflection on memory. Memory also serves the self. And when it serves the self, it is not only through its capacity to preserve, but also by what it makes possible in terms of recombination and reconstruction. It is its dynamic character that allows individuals to constitute themselves as persons.

In this sense, the memory theory that I have proposed allows us to think that people who remember inaccurately, or even confabulate, are also persons because they have this capacity to recognize themselves and thus to constitute themselves.

Bibliography

- Addis, Donna Rose, Katie Knapp, Reece P. Roberts, and Daniel L. Schacter. 2012. "Routes to the Past: Neural Substrates of Direct and Generative Autobiographical Memory Retrieval." *NeuroImage* 59(3):2908–22. doi: 10.1016/j.neuroimage.2011.09.066.
- Addis, Donna Rose, Ling Pan, Mai-Anh Vu, Noa Laiser, and Daniel L. Schacter. 2009. "Constructive Episodic Simulation of the Future and the Past: Distinct Subsystems of a Core Brain Network Mediate Imagining and Remembering." *Neuropsychologia* 47(11):2222–38. doi: 10.1016/j.neuropsychologia.2008.10.026.
- Addis, Donna Rose, Alana T. Wong, and Daniel L. Schacter. 2007. "Remembering the Past and Imagining the Future: Common and Distinct Neural Substrates during Event Construction and Elaboration." *Neuropsychologia* 45(7):1363–77. doi: 10.1016/j.neuropsychologia.2006.10.016.
- Anscombe, G. E. M. 1981. *Metaphysics and the Philosophy of Mind*. University of Minnesota Press.
- Atherton, Margaret. 1983. "Locke's Theory of Personal Identity." *Midwest Studies in Philosophy* 8(1):273–93.
- Barba, Gianfranco Dalla, Vincenzo Parlato, Antoinette Jobert, Yves Samson, and Sabina Pappatà. 1998. "Cortical Networks Implicated in Semantic and Episodic Memory: Common or Unique?" *Cortex*. doi: 10.1016/S0010-9452(08)70513-6.
- Barber, Kenneth F., and Jorge J. E. Gracia. 1994. *Individuation and Identity in Early Modern Philosophy: Descartes to Kant*. SUNY Press.

- Barner, Alma K. draft. "An Argument for Normative Discontinuism."
- Bartlett, F. C. 1932. *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press.
- Bartol, Curt R., and Anne M. Bartol. 2006. "History of Forensic Psychology." Pp. 3–27 in *The handbook of forensic psychology, 3rd ed.* Hoboken, NJ, US: John Wiley & Sons Inc.
- Beckett, Katherine. 1996. "Culture and the Politics of Signification: The Case of Child Sexual Abuse." *Social Problems* 43(1):57–76. doi: 10.1525/sp.1996.43.1.03x0336x.
- Berkeley, George. 1732. *Alciphron: Or, The Minute Philosopher. In Seven Dialogues. Containing an Apology for the Christian Religion Against Those Who Are Called Free-Thinkers ...* J. Tonson.
- Bernecker, Sven, ed. 2008. "The Motivation of the Causal Theory of Memory." Pp. 17–29 in *The Metaphysics of Memory, Philosophical Studies Series*. Dordrecht: Springer Netherlands.
- Bernecker, Sven. 2009. *Memory: A Philosophical Study*. Oxford, New York: Oxford University Press.
- Bernecker, Sven, Michaelian, Kourken eds., 2017. *The Routledge Handbook of Philosophy of Memory*. Routledge.
- Bernstein, Daniel M., and Elizabeth F. Loftus. 2009. "The Consequences of False Memories for Food Preferences and Choices." *Perspectives on Psychological Science* 4(2):135–39. doi: 10.1111/j.1745-6924.2009.01113.x.
- Blumen, Helena M., Suparna Rajaram, and Linda Henkel. 2013. "The Applied Value of Collaborative Memory Research in Aging: Behavioral and Neural Considerations." *Journal of Applied Research in Memory and Cognition* 2(2):107–17. doi: 10.1016/j.jarmac.2013.03.003.
- Boeker, Ruth. 2014. "The Moral Dimension in Locke's Account of Persons and Personal Identity." *History of Philosophy Quarterly* 31(3):229–247.
- Bortolotti, Lisa. 2015. "The Epistemic Innocence of Motivated Delusions." *Consciousness and Cognition* 33:490–99. doi: 10.1016/j.concog.2014.10.005.
- Bortolotti, Lisa, and Ema Sullivan-Bissett. 2018. "The Epistemic Innocence of Clinical Memory Distortions." *Mind & Language* 33(3):263–79. doi: 10.1111/mila.12175.
- Boyle, A. 2019. "Learning from the Past: Epistemic Generativity and the Function of Episodic Memory." *Journal of Consciousness Studies* 26(5–6):242–51.
- Braun, Kathryn A., Rhiannon Ellis, and Elizabeth F. Loftus. 2002. "Make My

- Memory: How Advertising Can Change Our Memories of the Past.” *Psychology and Marketing* 19(1):1–23. doi: 10.1002/mar.1000.
- Breen, N., D. Caine, and M. Coltheart. 2000. “Models of Face Recognition and Delusional Misidentification: A Critical Review.” *Cognitive Neuropsychology* 17(1):55–71. doi: 10.1080/026432900380481.
- Bruck, Maggie, Stephen J. Ceci, and Helene Hembrooke. 2002. “The Nature of Children’s True and False Narratives.” *Developmental Review* 22(3):520–54. doi: 10.1016/S0273-2297(02)00006-0.
- Burge, Tyler. 1989. “Individuation and Causation in Psychology.” *Pacific Philosophical Quarterly* 70: 303-322.
- Burnham, WH. 1904. “Retroactive Amnesia: Illustrative Cases and a Tentative Explanation.” *Am. J. Psychol.* 1904;14:382–396.
- Butler, Bishop. 1736. *The Works of Bishop Butler*. NED-New edition: 2006. edited by D. E. White. Boydell and Brewer.
- Calle, Sophie. 2013. *Ghosts*. Distributed Art Pub Incorporated.
- Campbell, Sue. 2014. *Our Faithfulness to the Past: The Ethics and Politics of Memory*. Oxford University Press.
- Candel, Ingrid, Harald Merckelbach, Katrijn Houben, and Inne Vandyck. 2004. “How Children Remember Neutral and Emotional Pictures: Boundary Extension in Children’s Scene Memories.” *The American Journal of Psychology* 117(2):249–57.
- Capgras, J., and J. Reboul-Lachaux. 1994. “L’Illusion Des ‘sosies’ Dans Un Délire Systématisé Chronique.” *History of Psychiatry* 5(17):119–33. doi: 10.1177/0957154X9400501709.
- Chadwick, Martin J., Sinéad L. Mullally, and Eleanor A. Maguire. 2013. “The Hippocampus Extrapolates beyond the View in Scenes: An fMRI Study of Boundary Extension.” *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior* 49(8):2067–79. doi: 10.1016/j.cortex.2012.11.010.
- Cheng, Sen, and Markus Werning. 2016. “What Is Episodic Memory If It Is a Natural Kind?” *Synthese* 193(5):1345–85. doi: 10.1007/s11229-014-0628-6.
- Chomsky, Noam. 1959. “Review of B.F. Skinner’s Verbal Behavior.” *Language* 35(1):26. doi: 10.2307/411334.
- Clark, Andrew, Robert A. Nash, Gabrielle Fincham, and Giuliana Mazzoni. 2012. “Creating Non-Believed Memories for Recent Autobiographical Events” edited by G. Andersson. *PLoS ONE* 7(3):e32998. doi: 10.1371/journal.pone.0032998.

- Clifford, BR, and J. Scott. 1978. "Individual and Situational Factors in Eyewitness Testimony." *JOURNAL OF APPLIED PSYCHOLOGY* Volume:63 Issue:3 Dated:(JUNE 1978) Pages:352-359.
- Coltheart, Max. 2017. "Confabulation and Conversation." *Cortex* 87:62–68. doi: 10.1016/j.cortex.2016.08.002.
- Conway, Martin A. 2005. "Memory and the Self." *Journal of Memory and Language* 53(4):594–628. doi: 10.1016/j.jml.2005.08.005.
- Conway, Martin A., and Catherine Loveday. 2015. "Remembering, Imagining, False Memories & Personal Meanings." *Consciousness and Cognition* 33:574–81. doi: 10.1016/j.concog.2014.12.002.
- Conway, Martin A., Jefferson A. Singer, and Angela Tagini. 2004. "The Self and Autobiographical Memory: Correspondence and Coherence." *Social Cognition* 22(5: Special issue):491–529.
- Copenhaver, Rebecca. 2006. "Thomas Reid's Theory of Memory." *History of Philosophy Quarterly* 23(2):171–89.
- Copenhaver, Rebecca. 2018. "Reid on Memory and Personal Identity." in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. Metaphysics Research Lab, Stanford University.
- Craver, Carl F. 2003. "The Making of a Memory Mechanism." *Journal of the History of Biology* 36(1):153–95. doi: 10.1023/A:1022596107834.
- Craver, Carl F. 2012. "A Preliminary Case for Amnesic Selves: Toward a Clinical Moral Psychology." *Social Cognition* 30(4):449–473.
- Craver, Carl F. 2020. "Remembering: Epistemic and Empirical." *Review of Philosophy and Psychology*. doi: 10.1007/s13164-020-00469-7.
- Crook, Lynn S. 1995. "Letter from Lynn Crook." *Journal of Child Sexual Abuse, Vol. 4(2)* 115–18.
- Crook, Lynn S., and Martha C. Dean. 1999. "Lost in a Shopping Mall. A Breach of Professional Ethics." *Ethics and Behavior* 9(1):39–50.
- De Brigard, Felipe. 2014. "Is Memory for Remembering? Recollection as a Form of Episodic Hypothetical Thinking." *Synthese* 191(2):155–85. doi: 10.1007/s11229-013-0247-7.
- Debus, Dorothea. 2008. "Experiencing the Past: A Relational Account of Recollective Memory." *Dialectica* 62(4):405–32. doi: 10.1111/j.1746-8361.2008.01165.x.
- Dennett, Daniel C. 1992. "The Self as a Center of Narrative Gravity." in *Self and Consciousness. Multiple Perspectives.*, edited by F. S. Kessel, P. M. Cole, and D. L. Johnson.

- DePrince, Anne P., Carolyn B. Allard, Hannah Oh, and Jennifer J. Freyd. 2004. "What's in a Name for Memory Errors? Implications and Ethical Issues Arising From the Use of the Term 'False Memory' for Errors in Memory for Details." *Ethics & Behavior* 14(3):201–33. doi: 10.1207/s15327019eb1403_1.
- Dewhurst, S., Craig Thorley, E. R. Hammond, and T. Ormerod. 2011. "Convergent, but Not Divergent, Thinking Predicts Susceptibility to Associative Memory Illusions." doi: 10.1016/J.PAID.2011.03.018.
- Dranseika, Vilius. 2020. "False Memories and Quasi-Memories Are Memories." in *Oxford Studies in Experimental Philosophy, Vol. 3.*, edited by Tania Lombrozo, Shaun Nichols & Joshua Knobe.
- Dussy, Dorothée. 2013. *Le berceau des dominations. Anthropologie de l'inceste - Dorothée Dussy*. La Discussion.
- Ebbinghaus, Hermann. 1885. *Memory: A Contribution to Experimental Psychology*. University Microfilms.
- Eustache, Francis, and Béatrice Desgranges. 2015. *Les Chemins de la mémoire*. Humensis.
- Fernández, Jordi. 2013. *Transparent Minds: A Study of Self-Knowledge*. Oxford University Press.
- Fernández, Jordi. 2016. "Epistemic Generation in Memory." *Philosophy and Phenomenological Research* 92(3):620–44. doi: 10.1111/phpr.12189.
- Flew, Antony. 1951. "Locke and the Problem of Personal Identity." *Philosophy* 26(96):53–68.
- Folescu, Marina. 2018. "Remembering Events: A Reidean Account of (Episodic) Memory." *Philosophy and Phenomenological Research* 97(2):304–21. doi: 10.1111/phpr.12333.
- Forest, Denis. 2018. "Who Am I? Lockean Questions for Contemporary Psychology and Brain Science." In Rodrigues P., Famerée J., eds., *The Genesis of Concepts and the Confrontation of Rationalities*, BETL (*Bibliotheca Ephemeridum Theologiarum Lovaniensium*), Peeters, Leuven, p 139-154.
- Forest, Denis, 2014. *Neurosepticisme. Les sciences du cerveau sous le scalpel de l'épistémologue*. Editions Ithaque.
- Forest, Denis, and Loraine Gérardin-Laverge. 2019. "L'homme Qui Prenait Sa Soeur Pour Une Étrangère." in Vol. Boissières (Fabienne), Jomand-Baudry (Régine). *L'énigme de la mémoire. Etudes pluridisciplinaires*, CNRS Editions, pp. 59–71, 2019, 978-2-271-09510-7.
- Fotopoulou, Aikaterini. 2008. "False Selves in Neuropsychological Rehabilitation: The Challenge of Confabulation." *Neuropsychological Rehabilitation* 18(5–6):541–65. doi: 10.1080/09602010802083545.

- Frazer, S. J., and J. M. Roberts. 1994. "Three Cases of Capgras' Syndrome." *The British Journal of Psychiatry: The Journal of Mental Science* 164(4):557–59. doi: 10.1192/bjp.164.4.557.
- Frey, U., Y. Y. Huang, and E. R. Kandel. 1993. "Effects of CAMP Simulate a Late Stage of LTP in Hippocampal CA1 Neurons." *Science (New York, N.Y.)* 260(5114):1661–64. doi: 10.1126/science.8389057.
- Garrett, Don. 2003. "Locke on Personal Identity, Consciousness, and" Fatal Errors"." *Philosophical Topics* 31(1/2):95–125.
- Garry, Maryanne, Charles G. Manning, Elizabeth F. Loftus, and Steven J. Sherman. 1996. "Imagination Inflation: Imagining a Childhood Event Inflates Confidence That It Occurred." *Psychonomic Bulletin & Review* 3(2):208–14. doi: 10.3758/BF03212420.
- Geach, P. T. 1967. "Identity." *Review of Metaphysics* 21(1):3–12. doi: revmetaph1967211196.
- Geraerts, Elke, Daniel M. Bernstein, Harald Merckelbach, Christel Linders, Linsey Raymaekers, and Elizabeth F. Loftus. 2008. "Lasting False Beliefs and Their Behavioral Consequences." *Psychological Science* 19(8):749–53. doi: 10.1111/j.1467-9280.2008.02151.x.
- Gérardin-Laverge, Loraine, and Denis Forest. 2018. "Forces et Faiblesses de La Théorie Narrative de l'identité Personnelle." in *Penser l'humain : définitions, descriptions, narrations.*, edited by M. Lequin.
- Gérardin-Laverge, Loraine, and Mona Gérardin-Laverge. 2015. "Multipistes. Récits de soi, mémoire, identité et langage." *Revue de neuropsychologie* me 7(2):69–70. doi: 10.1684/nrp.2015.0337.
- Gibson, James. 1917. *Locke's Theory of Knowledge and Its Historical Relations*. 2009th ed. Cornell University Library.
- Giustino, Thomas F., Paul J. Fitzgerald, and Stephen Maren. 2016. "Revisiting Propranolol and PTSD: Memory Erasure or Extinction Enhancement?" *Neurobiology of Learning and Memory* 130:26–33. doi: 10.1016/j.nlm.2016.01.009.
- Goldman, Alvin I. 1967. "A Causal Theory of Knowing." *The Journal of Philosophy* 64(12):357. doi: 10.2307/2024268.
- Gopnik, Alison. 2000. "Explanation as Orgasm and the Drive for Causal Knowledge: The Function, Evolution, and Phenomenology of the Theory Formation System." Pp. 299–323 in *Keil, J.C. and Wilson, R.A. (eds), Explanation and Cognition*.
- Gordon-Roth, Jessica. 2019. "Locke on Personal Identity." in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. Metaphysics Research Lab, Stanford University.

- Grant, Linda. 2011. *Remind Me Who I Am, Again*. Granta Books.
- Grice, H. P. 1941. "Personal Identity." *Mind* 50.
- Gustafsson, Johan E. 2010. "Did Locke Defend the Memory Continuity Criterion of Personal Identity?" *Locke Studies* 10:113–129.
- Gutchess, Angela H., and Daniel L. Schacter. 2012. "The Neural Correlates of Gist-Based True and False Recognition." doi: 10.1016/j.neuroimage.2011.11.078.
- Hamou, Philippe. 2004. "L'opinion de Locke sur la « matière pensante »." *Methodos. Savoirs et textes* (4). doi: 10.4000/methodos.123.
- Hamou, Philippe. 2014. "Mémoire et conscience continuée, une Lecture de Locke sur l'identité personnelle." *Philosophical enquiries : revue des philosophies anglophones* Dossier : Locke (II)(3):1–33.
- Hamou, Philippe, and Martine Pécharman, eds. 2018. *Locke and Cartesian Philosophy*. Oxford, New York: Oxford University Press.
- Harris, Celia B., Amanda J. Barnier, John Sutton, and Paul G. Keil. 2014. "Couples as Socially Distributed Cognitive Systems: Remembering in Everyday Social and Material Contexts." *Memory Studies*. doi: 10.1177/1750698014530619.
- Harris, Celia B., Amanda J. Barnier, John Sutton, and Greg Savage. 2019. "Features of Successful and Unsuccessful Collaborative Memory Conversations in Long-Married Couples." *Topics in Cognitive Science* 11(4):668–86. doi: 10.1111/tops.12350.
- Helm, Paul. 1979. "Locke's Theory of Personal Identity." *Philosophy* 54(208):173–85. doi: 10.1017/S0031819100048427.
- Henry, Jeremy, and Carl Craver. 2018. "Episodic Memory and the Witness Trump Card." *Behavioral and Brain Sciences* 41. doi: 10.1017/S0140525X17001376.
- Hirstein, William. 2005. *Brain Fiction: Self-Deception and the Riddle of Confabulation*. MIT Press.
- Hirstein, William, and V. S. Ramachandran. 1997. "Capgras Syndrome: A Novel Probe for Understanding the Neural Representation of the Identity and Familiarity of Persons." *Proceedings of the Royal Society of London B: Biological Sciences* 264(1380):437–44. doi: 10.1098/rspb.1997.0062.
- Hoerl, Christoph, and Teresa McCormack. 2001. *Time and Memory: Issues in Philosophy and Psychology*. Oxford University Press.
- Howe, Mark L., Sarah R. Garner, Monica Charlesworth, and Lauren Knott. 2011. "A Brighter Side to Memory Illusions: False Memories Prime Children's and Adults' Insight-Based Problem Solving." *Journal of Experimental*

Child Psychology 108(2):383–93. doi: 10.1016/j.jecp.2010.08.012.

Huang, Y. Y., and E. R. Kandel. 1994. “Recruitment of Long-Lasting and Protein Kinase A-Dependent Long-Term Potentiation in the CA1 Region of Hippocampus Requires Repeated Tetanization.” *Learning & Memory (Cold Spring Harbor, N.Y.)* 1(1):74–82.

Hutto, Daniel. 2017. “Memory and Narrativity.” *Faculty of Law, Humanities and the Arts - Papers (Archive)* 192–204.

Intraub, Helene. 2012. “Rethinking Visual Scene Perception.” *WIREs Cognitive Science* 3(1):117–27. doi: 10.1002/wcs.149.

Intraub, Helene, and Christopher Dickinson. 2008. “False Memory 1/20th of a Second Later: What the Early Onset of Boundary Extension Reveals about Perception.” *Psychological Science* 19(10):1007–14. doi: 10.1111/j.1467-9280.2008.02192.x.

Intraub, Helene, and Michael Richardson. 1989. “Wide-Angle Memories of Close-up Scenes.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15(2):179–87. doi: 10.1037/0278-7393.15.2.179.

Jajdelska, Elspeth, Miranda Anderson, Christopher Butler, Nigel Fabb, Elizabeth Finnigan, Ian Garwood, Stephen Kelly, Wendy Kirk, Karin Kukkonen, Sinead Mullally, and Stephan Schwan. 2019. “Picture This: A Review of Research Relating to Narrative Processing by Moving Image Versus Language.” *Frontiers in Psychology* 10. doi: 10.3389/fpsyg.2019.01161.

Johansson, N. O., J. Andersson, and J. Rönnerberg. 2005. “Compensating Strategies in Collaborative Remembering in Very Old Couples.” *Scandinavian Journal of Psychology*, 46(4), 349–359.

Johnson, Marcia K., Shahin Hashtroudi, and D. Stephen Lindsay. 1993. “Source Monitoring.” *Psychological Bulletin* 114(1):3–28. doi: 10.1037/0033-2909.114.1.3.

Kida, Satoshi. 2019. “Reconsolidation/Destabilization, Extinction and Forgetting of Fear Memory as Therapeutic Targets for PTSD.” *Psychopharmacology* 236(1):49–57. doi: 10.1007/s00213-018-5086-2.

Kopelman, M. D. 1987. “Two Types of Confabulation.” *Journal of Neurology, Neurosurgery, and Psychiatry* 50(11):1482–87.

Kopelman, Michael D. 2010. “Varieties of Confabulation and Delusion.” *Cognitive Neuropsychiatry* 15(1):14–37. doi: 10.1080/13546800902732830.

Kreindel, Erica, and Helene Intraub. 2017. “Anticipatory Scene Representation in Preschool Children’s Recall and Recognition Memory.” *Developmental Science*. doi: 10.1111/desc.12444.

La Corte, Valentina, Gianfranco Dalla Barba, Jean-Didier Lemaréchal, Line

- Garnero†, and Nathalie George. 2012. “Behavioural and Magnetoencephalographic Evidence for the Interaction Between Semantic and Episodic Memory in Healthy Elderly Subjects.” *Brain Topography* 25(4):408–22. doi: 10.1007/s10548-012-0222-5.
- Lackey, Jennifer. 2005. “Memory as a Generative Epistemic Source.” *Philosophy and Phenomenological Research* 70(3):636–58. doi: 10.1111/j.1933-1592.2005.tb00418.x.
- Laney, Cara, and Elizabeth Loftus. 2013. “Recent Advances in False Memory Research.” *South African Journal of Psychology* 43:137–46. doi: 10.1177/0081246313484236.
- Laney, Cara, and Elizabeth F. Loftus. 2008. “Emotional Content of True and False Memories.” *Memory (Hove, England)* 16(5):500–516. doi: 10.1080/09658210802065939.
- Lee, Henry. 1702. *Anti-Scepticism: Or, Notes Upon Each Chapter of Mr. Lock’s Essay Concerning Humane Understanding. With an Explication of All the Particulars of Which He Treats, and in the Same Order. In Four Books.* R. Clavel and C. Harper.
- Leibniz, Gottfried Wilhelm Freiherr von. 1765. *Leibniz: New Essays on Human Understanding.* Cambridge University Press.
- Lennon, Thomas M. 2001. “Locke and the Logic of Ideas.” *History of Philosophy Quarterly* 18(2):155–77.
- Letheby, Chris. 2016. “The Epistemic Innocence of Psychedelic States.” *Consciousness and Cognition* 39:28–37. doi: 10.1016/j.concog.2015.11.012.
- Lewis, David. 1976. “Survival and Identity.” in *The Identities of persons*, edited by A. Rorty. Berkeley : University of California Press.
- Locke, John. 1694. *An Essay Concerning Human Understanding.* London New-York: Dent Dutton, 1968.
- Locke, John, and John Sergeant. 1697. *Marginalia in John Sergeant’s Solid Philosophy.* Digital Locke Project.
- Loftus, E. F. 2005. “Planting Misinformation in the Human Mind: A 30-Year Investigation of the Malleability of Memory.” *Learning & Memory* 12(4):361–66. doi: 10.1101/lm.94705.
- Loftus, E. F., and Jacqueline E. Pickrell. 1995. “The Formation of False Memories.” *Psychiatric Annals* 25(12):720–25.
- Loftus, Elizabeth F. 1980. *Eyewitness Testimony.* Harvard University Press.
- Loftus, Elizabeth F. 1988. *Memory.* Rowman & Littlefield Publishers.

- Lømo, Terje. 2003. "The Discovery of Long-Term Potentiation." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 358(1432):617–20. doi: 10.1098/rstb.2002.1226.
- Lu, Yun-Fei, Eric R. Kandel, and Robert D. Hawkins. 1999. "Nitric Oxide Signaling Contributes to Late-Phase LTP and CREB Phosphorylation in the Hippocampus." *The Journal of Neuroscience* 19(23):10250–61. doi: 10.1523/JNEUROSCI.19-23-10250.1999.
- Lynch, M. A. 2004. "Long-Term Potentiation and Memory." *Physiological Reviews* 84(1):87–136. doi: 10.1152/physrev.00014.2003.
- MacIntyre, Alasdair. 1981. *After Virtue: A Study in Moral Theory*. University of Notre-Dame Press.
- Mackie, J. L. 1976. *Problems From Locke*. Clarendon Press.
- Maguire, Eleanor A., Helene Intraub, and Sinéad L. Mullally. 2016. "Scenes, Spaces, and Memory Traces: What Does the Hippocampus Do?" *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry* 22(5):432–39. doi: 10.1177/1073858415600389.
- Mahr, Johannes B., and Gergely Csibra. 2018. "Why Do We Remember? The Communicative Function of Episodic Memory." *Behavioral and Brain Sciences* 41. doi: 10.1017/S0140525X17000012.
- Mamus, Ezgi, and Aysecan Boduroglu. 2018. "The Role of Context on Boundary Extension." *Visual Cognition* 26(2):115–30. doi: 10.1080/13506285.2017.1399947.
- Martin, C. B., and Max Deutscher. 1966. "Remembering." *The Philosophical Review* 75(2):161–96. doi: 10.2307/2183082.
- Martin, M. G. F. 2015. "Old Acquaintance: Russell, Memory and Problems with Acquaintance." *Analytic Philosophy* 56(1):1–44. doi: 10.1111/phib.12059.
- Martin, Michael G. F. 2001. "Out of the Past: Episodic Recall as Retained Acquaintance." Pp. 257–284 in *Time and Memory*, edited by C. Hoerl and T. McCormack. Oxford University Press.
- Mazzoni, Giuliana, Alan Scoboria, and Lucy Harvey. 2010. "Nonbelieved Memories." *Psychological Science* 21(9):1334–40. doi: 10.1177/0956797610379865.
- McCarroll, Christopher. forthcoming. "Remembering the Personal Past: Beyond the Boundaries of Imagination."
- McCarroll, Christopher. 2018. *Remembering from the Outside: Personal Memory and the Perspectival Mind*. Oxford, New York: Oxford University Press.
- McCarroll, Christopher Jude. 2017. "Looking the Past in the Eye: Distortion in Memory and the Costs and Benefits of Recalling from an Observer

- Perspective.” *Consciousness and Cognition* 49:322–32. doi: 10.1016/j.concog.2017.01.014.
- McCarroll, Christopher Jude. 2019. “Looking at the Self: Perspectival Memory and Personal Identity.” *Philosophical Explorations* 22(3):259–79. doi: 10.1080/13869795.2018.1562087.
- McCarroll, Christopher Jude, and John Sutton. 2017. “Memory and Perspective.” P. 25 in Vol. *Routledge Handbook of Philosophy of Memory*, edited by S. Bernecker and K. Michaelian.
- McClelland, James L. 1995. “Constructive Memory and Memory Distortions: A Parallel-Distributed Processing Approach.” P. 12 in Vol. *Memory Distortion*. Schacter (Ed).
- McDunn, Benjamin A., Aisha P. Siddiqui, and James M. Brown. 2014. “Seeking the Boundary of Boundary Extension.” *Psychonomic Bulletin & Review* 21(2):370–75. doi: 10.3758/s13423-013-0494-0.
- McNally, Richard J., Natasha B. Lasko, Susan A. Clancy, Michael L. Macklin, Roger K. Pitman, and Scott P. Orr. 2004. “Psychophysiological Responding During Script-Driven Imagery in People Reporting Abduction by Space Aliens.” *Psychological Science* 15(7):493–97. doi: 10.1111/j.0956-7976.2004.00707.x.
- Meade, Michelle L., Celia B. Harris, Penny Van Bergen, John Sutton, and Amanda J. Barnier, eds. 2017. *Collaborative Remembering: Theories, Research, and Applications*. Oxford, New York: Oxford University Press.
- Merrill, Nathalie, Theodore E. Waters, and Robyn Fivush. 2015. “Connecting the Self to Traumatic and Positive Events: Links to Identity and Well-Being.” *Memory* 24(10):1-8.
- Michaelian, Kourken. 2011. “Generative Memory.” *Philosophical Psychology* 24(3):323–42. doi: 10.1080/09515089.2011.559623.
- Michaelian, Kourken. 2013. “The Information Effect: Constructive Memory, Testimony, and Epistemic Luck.” *Synthese* 190:1–28. doi: 10.1007/s11229-011-9992-7.
- Michaelian, Kourken. 2016a. “Against Discontinuism: Mental Time Travel and Our Knowledge of Past and Future Events.” *Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel*.
- Michaelian, Kourken. 2016b. “Confabulating, Misremembering, Relearning: The Simulation Theory of Memory and Unsuccessful Remembering.” *Frontiers in Psychology* 7:1857. doi: 10.3389/fpsyg.2016.01857.
- Michaelian, Kourken. 2016c. *Mental Time Travel: Episodic Memory and Our Knowledge of the Personal Past*.
- Michaelian, Kourken. 2018. “Confabulating as Unreliable Imagining: In Defence

- of the Simulationist Account of Unsuccessful Remembering.” *Topoi*. doi: 10.1007/s11245-018-9591-z.
- Michaelian, Kourken, and Sarah Robins. 2018. “Beyond the Causal Theory? Fifty Years After Martin and Deutscher.” Pp. 13–32 in *New Directions in the Philosophy of Memory*, edited by K. Michaelian, D. Debus, and D. Perrin. Routledge.
- Michaelian, Kourken, and André Sant’Anna. forthcoming. “From Authenticism to Alethism: Against McCarroll on Observer Memory.”
- Millikan, Ruth G. 1984. *Language, Thought, and Other Biological Categories*. MIT Press.
- Moscovitch, Morris. 1995. “Confabulation.” Pp. 226–51 in *Memory distortions: How minds, brains, and societies reconstruct the past*. Cambridge, MA, US: Harvard University Press.
- Mullally, Sinéad L., Helene Intraub, and Eleanor A. Maguire. 2012. “Attenuated Boundary Extension Produces a Paradoxical Memory Advantage in Amnesic Patients.” *Current Biology* 22(4):261–68. doi: 10.1016/j.cub.2012.01.001.
- Mullally, Sinéad L., Faraneh Vargha-Khadem, and Eleanor A. Maguire. 2014. “Scene Construction in Developmental Amnesia: An fMRI Study.” *Neuropsychologia* 52:1–10. doi: 10.1016/j.neuropsychologia.2013.11.001.
- Mulligan, Neil W., and Susan L. Hornstein. 2003. “Memory for Actions: Self-Performed Tasks and the Reenactment Effect.” *Memory & Cognition* 31(3):412–21. doi: 10.3758/BF03194399.
- Münsterberg, Hugo. 1908. *On the Witness Stand: Essays on Psychology and Crime*.
- Nadel, Lynn, and Morris Moscovitch. 1997. “Memory Consolidation, Retrograde Amnesia and the Hippocampal Complex.” *Current Opinion in Neurobiology Volume 7, Issue 2, April 1997, Pages 217-227* 11.
- Nanay, Bence. 2010. “A Modal Theory of Function.” *Journal of Philosophy* 107(8):412–431. doi: 10.5840/jphil2010107834.
- Nanay, Bence. 2012. “Function Attribution Depends on the Explanatory Context: A Reply to Neander and Rosenberg’s Reply to Nanay.” *Journal of Philosophy* 109(10):623–627. doi: 10.5840/jphil20121091031.
- Nanay, Bence. 2018a. *Seeing Things You Don’t See*. Oxford University Press. Oxford.
- Nanay, Bence. 2018b. “The Importance of Amodal Completion in Everyday Perception.” *I-Perception* 9(4):2041669518788887. doi: 10.1177/2041669518788887.

- Neander, Karen. 1991. "Types of Traits: The Importance of Functional Homologues." in *Functions: New Essays in the Philosophy of Psychology and Biology*, edited by A. Ariew, R. Cummins, and M. Perlman.
- Neisser, Ulric. 1988. "Five Kinds of Self-Knowledge." *Philosophical Psychology* 1(1):35–59.
- Newman, Lex. 2015. "Locke on Substance, Consciousness, and Personal Identity." Pp. 128–145. in, edited by Lodge and Stoneham.
- O'Donnell, Colin, Gary Donohoe, Louise Sharkey, Nicholas Owens, Maria Migone, Raewynn Harries, Anthony Kinsella, Conall Larkin, and Eadbhard O'Callaghan. 2003. "Compliance Therapy: A Randomised Controlled Trial in Schizophrenia." *BMJ: British Medical Journal* 327(7419):834.
- Olson, Eric T. 2017. "Personal Identity." in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. Metaphysics Research Lab, Stanford University.
- Otgaar, Henry, Alan Scoboria, and Giuliana Mazzoni. 2014. "On the Existence and Implications of Nonbelieved Memories." *Current Directions in Psychological Science* 23(5):349–54. doi: 10.1177/0963721414542102.
- Otgaar, Henry, Alan Scoboria, and Tom Smeets. 2012. "Experimentally Evoking Nonbelieved Memories for Childhood Events." *Journal of Experimental Psychology. Learning, Memory, and Cognition* 39:717–30. doi: 10.1037/a002966839.
- Ott, Walter R. 2003. *Locke's Philosophy of Language*. Cambridge: Cambridge University Press.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford, New York: Oxford University Press.
- Perrin, Denis. 2016. "Asymmetries in Subjective Time." in *Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel*. Eds. K. Michaelian, S. B. Klein, and K. K. Szpunar. New York: Oxford University Press.
- Perrin, Denis. 2017. "Souvenir (A)." *L'Encyclopédie Philosophique*. Retrieved October 19, 2018 (<https://encyclo-philo.fr/souvenir-a/>).
- Perrin, Denis, and Kourken Michaelian. 2017. "MEMORY AS MENTAL TIME TRAVEL." 12.
- Perry, David L. 1967. "Locke on Mixed Modes, Relations, and Knowledge." *Journal of the History of Philosophy* 5(3):219–235. doi: 10.1353/hph.2008.1493.
- Perry, John, ed. 2008. *Personal Identity*. 2 edition. Berkeley, Calif. London: University of California Press.

- Perry, John. 2017. "Identity and Self-Knowledge." *Philosophy, Theory, and Practice in Biology* 9(20171201). doi: 10.3998/ptb.6959004.0009.005.
- Quinn, Paul C., and Helene Intraub. 2007. "Perceiving 'Outside the Box' Occurs Early in Development: Evidence for Boundary Extension in Three- to Seven-Month-Old Infants." *Child Development* 78(1):324–34. doi: 10.1111/j.1467-8624.2007.01000.x.
- Quinton, Anthony. 1962. "The Soul." *Journal of Philosophy* 59(15):393–409. doi: 10.2307/2022957.
- Ramachandran, V. S. 1998. "Consciousness and Body Image: Lessons from Phantom Limbs, Capgras Syndrome and Pain Asymbolia." *Philosophical Transactions of the Royal Society B: Biological Sciences* 353(1377):1851–59.
- Reid, Thomas. 1785. *Essays on the Intellectual Powers of Man*. 1852nd ed. J. Bartlett.
- Rickless, Samuel C. 2018. "Locke's Ontology of Relations." *Locke Studies* 17:61–86. doi: 10.5206/ls.2017.870.
- Ricoeur, Paul. 1992. *Oneself as Another*. University of Chicago Press.
- Robins, Sarah. forthcoming. "No Reliability in Remembering."
- Robins, Sarah. 2020. "Defending Discontinuism, Naturally." *Review of Philosophy and Psychology* 11(2):469–86. doi: 10.1007/s13164-020-00462-0.
- Robins, Sarah. 2017. "Confabulation and Constructive Memory." *Synthese*. doi: 10.1007/s11229-017-1315-1.
- Robins, Sarah. 2016a. "Misremembering." *Philosophical Psychology* 29(3):432–47. doi: 10.1080/09515089.2015.1113245.
- Robins, Sarah. 2016b. "Representing the Past: Memory Traces and the Causal Theory of Memory." *Philosophical Studies* 173(11):2993–3013. doi: 10.1007/s11098-016-0647-x.
- Roediger, Henry L., and Kathleen B. McDermott. 1995. "Creating False Memories: Remembering Words Not Presented in Lists." 12.
- Rosenthal, David M. 2009. "Higher-Order Theories of Consciousness." *The Oxford Handbook of Philosophy of Mind*. Retrieved September 27, 2020 (<https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199262626.18.001.0001/oxfordhb-9780199262618-e-14>).
- Saad, Brad. forthcoming. "Reflections in Mind: Consciousness and Time Reversal Invariance."
- Sant'Anna, André. forthcoming. "Relationalism and the Transparency of

Recollective Memory.”

- Sant’Anna, André. forthcoming. “Unsuccessful Remembering: A Challenge for the Relational View of Memory.” *Erkenntnis* 1–24. doi: 10.1007/s10670-020-00261-0.
- Sant’Anna, André. 2018. “The Hybrid Contents of Memory.” *Synthese*. doi: 10.1007/s11229-018-1753-4.
- Sarason, I. G., and R. Stoops. 1978. “Test Anxiety and the Passage of Time.” *Journal of Consulting and Clinical Psychology*, 46(1), 102–109.
- Schacter, Daniel L. 1997. *Searching For Memory: The Brain, The Mind, And The Past* by Daniel L. Schacter. Basic Books.
- Schacter, Daniel L. 2008. *Searching For Memory: The Brain, The Mind, And The Past*. Hachette UK.
- Schacter, Daniel L., and Donna Rose Addis. 2007. “Constructive Memory: The Ghosts of Past and Future.” *Nature* 445(7123):27–27.
- Schacter, Daniel L, and Donna Rose Addis. 2007. “The Cognitive Neuroscience of Constructive Memory: Remembering the Past and Imagining the Future.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 362(1481):773–86. doi: 10.1098/rstb.2007.2087.
- Schacter, Daniel L., Donna Rose Addis, and Randy L. Buckner. 2007. “Remembering the Past to Imagine the Future: The Prospective Brain.” *Nature Reviews Neuroscience* 8(9):657–661.
- Schacter, Daniel L., Scott A. Guerin, and Peggy L. St. Jacques. 2011. “Memory Distortion: An Adaptive Perspective.” *Trends in Cognitive Sciences* 15(10):467–74. doi: 10.1016/j.tics.2011.08.004.
- Schacter, Daniel L., and Elizabeth F. Loftus. 2013. “Memory and Law: What Can Cognitive Neuroscience Contribute?” 16(2):119–23. doi: 10.1038/nn.3294.
- Schechtman, Marya. 1994. “The Truth about Memory.” *Philosophical Psychology* 7(1):3–18. doi: 10.1080/09515089408573107.
- Schechtman, Marya. 1996. *The Constitution of Selves*. Cornell University Press.
- Schechtman, Marya. 2001. “Empathic Access: The Missing Ingredient in Personal Identity.” *Philosophical Explorations* 4(2):95–111. doi: 10.1080/10002001058538710.
- Schechtman, Marya. 2016. “Personal Identity.” in *Routledge Encyclopedia of Philosophy*. London: Routledge.
- Schnider, A., C. von Daniken, and K. Gutbrod. 1996. “The Mechanisms of Spontaneous and Provoked Confabulations.” *Brain* 119(4):1365–75. doi:

10.1093/brain/119.4.1365.

- Schulz, Kathryn. 2010. *Being Wrong. Adventures in The Margin of Error*. Ecco Press.
- Scoboria, Alan, Giuliana Mazzoni, Josée L. Jarry, and Daniel M. Bernstein. 2012. "Personalized and Not General Suggestion Produces False Autobiographical Memories and Suggestion-Consistent Behavior." *Acta Psychologica* 139(1):225–32. doi: 10.1016/j.actpsy.2011.10.008.
- Seamon, John G., Morgan M. Philbin, and Liza G. Harrison. 2006. "Do You Remember Proposing Marriage to the Pepsi Machine? False Recollections from a Campus Walk." *Psychonomic Bulletin & Review* 13(5):752–56.
- Seamon, John G., Sarah E. Schlegel, Peter M. Hiester, Susan M. Landau, and Brianne F. Blumenthal. 2002. "Misremembering Pictured Objects: People of All Ages Demonstrate the Boundary Extension Illusion." *The American Journal of Psychology* 115(2):151–67.
- Sellnow, Deanna D., and Kristen P. Treinen. 2004. "The Role of Gender in Perceived Speaker Competence: An Analysis of Student Peer Critiques." *Communication Education* 53(3):286–96. doi: 10.1080/0363452042000265215.
- Sergeant, John. 1697. *Solid Philosophy Asserted, against the Fancies of the Ideists, or, The Method to Science Farther Illustrated with Reflexions on Mr. Locke's Essay Concerning Human Understanding / by J.S.* London: Printed for Roger Clavil ... Abel Roper ... and Thomas Metcalf.
- Shallice, T. 1997. "Modularity and Consciousness." in N. Block, O. Flanagan, & G. Guzeldere (Eds.), *The nature of consciousness*. Cambridge: MIT Press.
- Shaw, Julia, and Stephen Porter. 2015. "Constructing Rich False Memories of Committing Crime." *Psychological Science* 26(3):291–301. doi: 10.1177/0956797614562862.
- Shoemaker, Sydney. 1970. "Persons and Their Pasts." *American Philosophical Quarterly* 7(4):269–85.
- Skinner, B. F. 1957. *Verbal Behavior*. 2015 reprint of 1957 Edition. Mansfield Centre: Martino Fine Books.
- Soteriou, Matthew. 2000. "The Particularity of Visual Perception." *European Journal of Philosophy* 8(2):173–189. doi: 10.1111/1468-0378.00107.
- Squire, L. R. 1992. "Memory and the Hippocampus: A Synthesis from Findings with Rats, Monkeys, and Humans." *Psychological Review* 99(2):195–231. doi: 10.1037/0033-295x.99.2.195.
- Strawson, Galen. 2004. "Against Narrativity." *Ratio* 17(4):428–452.
- Strawson, Galen. 2011. *Locke on Personal Identity: Consciousness and*

- Concernment*. Princeton University Press.
- Stuart, Matthew. 2015. *A Companion to Locke*. John Wiley & Sons.
- Sullivan-Bissett, Ema. 2015. "Implicit Bias, Confabulation, and Epistemic Innocence." *Consciousness and Cognition* 33:548–560.
- Sutton, John. 2008. *Philosophy and Memory Traces: Descartes to Connectionism*. Cambridge University Press.
- Tabb, Kathryn. 2018. "Madness as Method: On Locke's Thought Experiments about Personal Identity." *British Journal for the History of Philosophy* 26(5):871–89. doi: 10.1080/09608788.2017.1350936.
- Tanaka, Kazumasa Z., and Thomas J. McHugh. 2018. "The Hippocampal Engram as a Memory Index." *Journal of Experimental Neuroscience* 12. doi: 10.1177/1179069518815942.
- Taylor, Charles. 1989. *Sources of the Self: The Making of the Modern Identity*. Cambridge University Press.
- Thiel, Udo. 2011. *The Early Modern Subject: Self-Consciousness and Personal Identity from Descartes to Hume*. OUP Oxford.
- Thiel, Udo. 2015. "Self and Sensibility: From Locke to Condillac and Rousseau." *Intellectual History Review* 25(3):257–78. doi: 10.1080/17496977.2014.888615.
- Tulving, Endel. 1972. "Episodic and Semantic Memory." in *E Tulving W Donaldson Eds from CED 529 at Arizona State University*.
- Tulving, Endel. 1984. "Précis of Elements of Episodic Memory." *Behavioral and Brain Sciences* 7(2):223. doi: 10.1017/s0140525x0004440x.
- Tulving, Endel. 1985. "Memory and Consciousness." *Canadian Psychology* 26:1–12.
- Tulving, Endel. 2002. "Episodic Memory: From Mind to Brain." *Annual Review of Psychology* 53(1):1–25.
- Urgolites, Zhisen J., Christine N. Smith, and Larry R. Squire. 2015. "True and False Memories, Parietal Cortex, and Confidence Judgments." *Learning & Memory* 22(11):557–62. doi: 10.1101/lm.038349.115.
- Velleman, J. David. 1996. "Self to Self." *The Philosophical Review* 105(1):39–76. doi: 10.2307/2185763.
- Velleman, J. David, ed. 2005. "The Self as Narrator." Pp. 203–23 in *Self to Self: Selected Essays*. Cambridge: Cambridge University Press.
- Viera, Gerardo. unpublished. "Time, Memory and the Self."

- Vieweg, W. Victor R., Demetrios A. Julius, Antony Fernandez, Mary Beatty-Brooks, John M. Hetteema, and Anand K. Pandurangi. 2006. "Posttraumatic Stress Disorder: Clinical Features, Pathophysiology, and Treatment." *The American Journal of Medicine* 119(5):383–90. doi: 10.1016/j.amjmed.2005.09.027.
- Waters, Theodore E. A., and Robyn Fivush. 2015. "Relations Between Narrative Coherence, Identity, and Psychological Well-Being in Emerging Adulthood." *Journal of Personality* 83(4):441–51. doi: 10.1111/jopy.12120.
- Weinberg, Shelley. 2011. "Locke on Personal Identity: Locke on Personal Identity." *Philosophy Compass* 6(6):398–407. doi: 10.1111/j.1747-9991.2011.00402.x.
- Weinberg, Shelley. 2012a. "The Metaphysical Fact of Consciousness in Locke's Theory of Personal Identity." *Journal of the History of Philosophy* 50(3):387–415.
- Weinberg, Shelley. 2012b. "The Metaphysical Fact of Consciousness in Locke's Theory of Personal Identity." *Journal of the History of Philosophy* 50(3):387–415.
- Weinberg, Shelley. 2016. "Locke on Knowing Our Own Ideas (and Ourselves)." *Pacific Philosophical Quarterly* 97(3):347–70. doi: 10.1111/papq.12066.
- Wiggins, David. 1967. *Identity and Spatio-Temporal Continuity*. Blackwell Oxford.
- Williams, Bernard. 1973. "Problems of the Self by Bernard Williams." *Cambridge Core*. Retrieved November 14, 2018 (/core/books/problems-of-the-self/440C25B26D0E943C22DE8C2B6186F4D9).
- Wilson, Timothy de Camp, and Richard E. Nisbett. 1978. "The Accuracy of Verbal Reports About the Effects of Stimuli on Evaluations and Behavior." *Social Psychology* 41(2):118–31. doi: 10.2307/3033572.
- Wiltgen, Brian J., and Alcino J. Silva. 2007. "Memory for Context Becomes Less Specific with Time." *Learning & Memory (Cold Spring Harbor, N.Y.)* 14(4):313–17. doi: 10.1101/lm.430907.
- Winkler, Kenneth. 1991. "Locke on Personal Identity." *Journal of the History of Philosophy* 29(2):201–226.
- Winocur, Gordon, and Morris Moscovitch. 2011. "Memory Transformation and Systems Consolidation." *Journal of the International Neuropsychological Society: JINS* 17(5):766–80. doi: 10.1017/S1355617711000683.
- Winocur, Gordon, Morris Moscovitch, and Bruno Bontempi. 2010. "Memory Formation and Long-Term Retention in Humans and Animals: Convergence towards a Transformation Account of Hippocampal-Neocortical Interactions." *Neuropsychologia* 48(8):2339–56. doi:

10.1016/j.neuropsychologia.2010.04.016.

- Yaffe, Gideon. 2007. "Locke on Ideas of Identity and Diversity." Pp. 192–230 in *The Cambridge Companion to Locke's "Essay Concerning Human Understanding,"* edited by L. Newman. Cambridge University Press.
- Yolton, John W. 2000. *Realism and Appearances: An Essay in Ontology.* Cambridge University Press.
- Yolton, John W. 2004. *The Two Intellectual Worlds of John Locke: Man, Person, and Spirits in the "Essay."* Cornell University Press.
- Zangrilli, Alessia, Giuseppe Ducci, Pier Luca Bandinelli, Jemima Dooley, Rosemarie McCabe, and Stefan Priebe. 2014. "How Do Psychiatrists Address Delusions in First Meetings in Acute Care? A Qualitative Study." *BMC Psychiatry* 14(1):178. doi: 10.1186/1471-244X-14-178.
- Zhu, Bi, Chuansheng Chen, Elizabeth F. Loftus, Qinghua He, Chunhui Chen, Xuemei Lei, Chongde Lin, and Qi Dong. 2013. *Brief Exposure to Misinformation Can Lead to Long-Term False Memories.* SSRN Scholarly Paper. ID 2201939. Rochester, NY: Social Science Research Network.

Full table of contents

RESUME	4
ABSTRACT	5
AFFILIATIONS	7
RESUME SUBSTANTIEL.....	9
PREMIERE PARTIE	10
PREMIER CHAPITRE	10
SECOND CHAPITRE	14
SECONDE PARTIE	18
TROISIEME CHAPITRE	19
QUATRIEME CHAPITRE.....	23
TROISIEME PARTIE	27
CINQUIEME CHAPITRE	28
CONCLUSION	33
ACKNOWLEDGMENTS.....	36
TABLE OF CONTENTS.....	43
INTRODUCTION	47
SITUATION	47
PROBLEMS AND DIRECTIONS	49
METHOD	50
OUTLINE.....	50
FIRST PART – MEMORY AND PERSONAL IDENTITY	52

CHAPTER 1 – AGAINST CIRCULARITY.....	53
INTRODUCTION	53
SECTION 1 – THE THREE CLAIMS.....	55
1.1. Epistemic claim	56
1.2. Reduction claim.....	61
1.3. Metaphysical claim	62
1.3.1. Consciousness and/or memory?.....	62
1.3.2. An objective fact?	65
SECTION 2 – THE CIRCULARITY OBJECTION	67
SECTION 3 – AGAINST CIRCULARITY: SELF-IDENTIFICATION WITHOUT PRIOR PERSONAL IDENTITY	70
3. 1. Building the question raised by the circularity objection	71
3.2 Ontology of relations in Locke	71
3.3. Fatal error or self-identification without prior personal identity.....	73
3.3.1. Transfer of consciousness?	74
3.3.2. First person perspective.....	74
3.3.3. False memory phenomenology	76
3.3.4. Objections	76
SECTION 4 – CONCLUDING REMARKS.	78
CHAPTER 2 – AGAINST THE PRESERVATIVE VIEW OF MEMORY	81
INTRODUCTION	81
SECTION 1 – LOCKE AND MEMORY AS A STOREHOUSE	84
1.1. Memory relation of identity	84
1.2. Remembering and the preservative view	85
SECTION 2 – BERKELEY’S AND REID’S OBJECTIONS.....	93
2.1. Reid’s exemplification of the transitivity problem.	93
2.2. Against the preservative view.....	95
SECTION 3 – LOCKEAN RESPONSES	97
3.1. Adapting the concept of memory.	98
3.2. De-emphasizing the role of memory: consciousness is the key ingredient of personal identity	101
SECTION 4 – A HYBRID VIEW OF EPISODIC MEMORY.....	107
4.1. First, the witness condition does not secure the sameness condition	109
4.1.1. Eyewitness testimony and memory distortions.	109
4.1.2. Constructive encoding	110
4.1.3. Reconstruction in Repeated Reproduction.....	112
4.2. Memory storage.....	114
4.3. A hybrid view of memory.....	116
4.3.1. Storage, consolidation and semantization	116
4.3.2. The adaptive dimension of memory distortions.....	117
SECTION 5 – BACK TO THE OBJECTION	119
SECTION 6 – CONCLUDING REMARKS	120

SECOND PART – THE CONSTRUCTIVE DIMENSION OF EPISODIC MEMORY124

CHAPTER 3 – ARE FALSE MEMORIES MEMORIES? THE CONDITIONS FOR REMEMBERING.	125
INTRODUCTION	125
SECTION 1 – A CASE.....	127
SECTION 2 – THE FACTIVITY CONDITION AS A NECESSARY CONDITION FOR REMEMBERING.....	132
2.1. The analogy between fake guns, false money and false memory.....	132
2.2. The factivity condition	138
2.3. Disjunctivism	139

2.3.1. Between genuine memories and memories of hallucinations	139
2.3.2. Between genuine and false memories.....	140
i. Phenomenal plan.....	141
ii. Neural plan	141
2.4. Factive parts	142
2.5. Misremembering.....	142
2.6. What does it mean that a false memory is false?	143
2.6.1. A problem with the term 'false memory'	143
2.6.2. Typical cases of false memories in psychology as results of a misinformation effect.	144
i. Bugs Bunny case (Braun, Ellis, and Loftus 2002).	144
ii. Lost In the Mall (LIM) case (Loftus and Pickrell 1995).....	144
2.7. Disentangle post-event information integration and falsity.....	146
2.7.1. Representations of a past event caused by the past experience can be false.	147
2.7.2. Integration of post-event information into a memory can lead to an accurate (true) representation of the past event.	148
i. Information effect.	148
ii. Epistemic contexts of the integration of the new piece.	148
2.8. Concluding remarks on factivity.....	151
SECTION 3 – THE CAUSAL CONDITION AS A NECESSARY CONDITION FOR REMEMBERING	152
3.1. The causal theory of memory	153
3.2. The boundaries of the past experience.....	157
3.2.1. Remembering more – inattentive encoding	157
3.2.2. Remembering more – filling gaps	159
3.3. The past experience as the source -- Source monitoring error	159
3.3.1. The phenomenon of boundary extension and the gathering of information from various sources at encoding.....	160
3.3.2. Interdependence of episodic and semantic memory at retrieval and conclusion on the causal condition.	163
3.4. Concluding remarks on the causal condition.	163
SECTION 4 – CONCLUDING REMARKS	165
CHAPTER 4 – NON-EPISTEMIC MEMORY EVALUATIONS AND NON-EPISTEMIC MEMORY FUNCTIONS.167	
INTRODUCTION.....	167
SECTION 1 – EPISTEMIC MEMORY AND EMPIRICAL MEMORY AS DISTINCT	168
1.1. Causal relations	169
1.2. First-hand experience.	170
1.3. Normative discontinuism	171
SECTION 2 – EPISTEMIC EVALUATIONS OF EPISODIC MEMORY DEPEND ON CONTEXTS	172
2.1. No alternative condition and epistemic innocence	172
2.2. Epistemic benefits	173
SECTION 3 – MEMORY SCIENCE LEADS TO REVISING EPISTEMIC MEMORY EVALUATIONS.	175
3.1. Clinical memory distortions and non-clinical memory distortions.	175
3.2. Interactions between memory science and memory evaluations.....	177
3.3. How to know the past on the basis of memory reports? Information and interpretation.....	178
SECTION 4 – NON-EPISTEMIC FUNCTIONS OF EPISODIC MEMORY	181
4.1. Contextualism about function attribution	181
4.2. Beneficial vs/ Unbeneficial.....	182
4.2.1. Unbeneficial accurate memories.	182
4.2.2. Beneficial memory distortions and psychological benefits of clinical confabulations	183
4.3. Inhibitory vs/ Empowering.....	183

SECTION 5 – CONCLUDING REMARKS	185
--------------------------------------	-----

THIRD PART – RETHINKING MEMORY TO REDEFINE PERSONAL IDENTITY.....188

CHAPTER 5 – PERSONAL IDENTITY, CONSTRUCTIVE MEMORY AND NARRATIVITY.	189
INTRODUCTION	189
SECTION 1 – THE MISSING PIECES OF MEMORY THEORIES	191
1.1. Construction.....	191
1.2. Social dimension	192
1.3. Envisioning one’s future.....	193
SECTION 2 – MACINTYRE’S NARRATIVE VIEW.	195
2.1. Construction of a coherent representation of myself as the foundation of personal identity.....	196
2.2. Frames of self-narratives, and the concept of embedment. Individualistic worry.	196
2.3. The meaningless life worry vs intelligibility – narratives and intentions.....	197
SECTION 3 – INCOMPLETENESS OF MEMORY AND NARRATIVE THEORIES – THEIR COMMON PRESUPPOSITIONS.....	199
SECTION 4 – AGAINST NARRATIVITY.....	200
4.1. Episodic selves.	201
4.2. Coherence	202
4.3. Personal identity without narrativity. Non-declarative episodic memory.	205
SECTION 5 – AN ATTEMPT TO REVISE THE MEMORY THEORY OF PERSONAL IDENTITY: A CONSTRUCTIVE MEMORY THEORY.	206
5.1. Envisioning the future.....	207
5.2. Shared-memory space – myself and others	208
5.3. Constructive memory theory of personal identity.	210
5.3.1. Epistemic and metaphysical questions.....	210
5.3.2. Circularity again?.....	211
5.3.3. What about self-deception?	211
5.3.4. What about extreme cases of self-deception such as in confabulation?	211
5.3.5. Amnesic selves.	212
SECTION 6 – CONCLUDING REMARKS.	214

CONCLUSION217

MEMORY AND PERSONAL IDENTITY IN LOCKE	218
PERSONAL IDENTITY AND CIRCULARITY	218
MEMORY AND THE PRESERVATIVE VIEW	221
THE CONSTRUCTIVE DIMENSION OF EPISODIC MEMORY	226
THE CONDITIONS FOR REMEMBERING.....	226
NON-EPISTEMIC FUNCTIONS OF EPISODIC MEMORY	230
CONSTRUCTIVE MEMORY AND PERSONAL IDENTITY	234

BIBLIOGRAPHY.....242

FULL TABLE OF CONTENTS.....262

