On the leaks and boundaries of imagination

Andrea Rivadulla Duró

UNIVERSITAT DE BARCELONA

On the leaks and boundaries of imagination

Andrea Rivadulla Duró

On the leaks and boundaries of imagination

Doctoral thesis submitted by Andrea Rivadulla Duró to the University of Barcelona for the degree of Doctor of Philosophy

Supervisors:
Manuel García-Carpintero Sánchez-Miguel
and Josep Macià Fàbrega

Ph.D. Program: Ciència Cognitiva i Llenguatge
Research line: Filosofia analítica

Facultat de Filosofia
Universitat de Barcelona
June 2022

A mis padres, hermano y hermanas
A Marc

Abstract

The present dissertation explores the role of imagination in different philosophical domains of inquiry. The thesis set out as a collection of self-standing essays and can be divided into two parts: *leaks of imagination* and *boundaries of imagination*. The first part concerns what I name *leaks of imagination*: Effects imagination has on attitudes and behavior. In Chapter 1, I review clinical and empirical evidence on the consequences imagining experiences has. To account for this evidence, I propose a theory—the *Prima Facie* View—and argue for the implicit assertoric force of imagination. According to this view, experiential imagination is not epistemically innocuous. Chapter 2 concerns the role of imagination in intrinsic symbolic actions. I argue that these actions elude an explanation in terms of a belief-desire pair or an emotion, and characterize them as symbolically displaced imaginings. The second part of the thesis, *boundaries of imagination,* delimits the appeal to the imagination. Chapter 3 criticizes the appeal to the imagination to explain the functional profile of delusions. This criticism is followed by a positive doxastic account of delusions in a fragmented and psychofunctional system of belief. Chapter 4 examines the Simulation Theory of Memory, which reduces episodic memory to imagination. I argue that given the way it equates episodic memory with imagination, the theory is in a compromised position to account for the characteristic phenomenology of episodic memory, and thus for its reliability. Overall, in this dissertation, I present a critical overview of four independent fields and propose new accounts of the problems at hand.

Acknowledgments *(Agradecimientos)*

I am immensely grateful to many people who have helped and accompanied me during the Ph.D. I want to thank my supervisor, Manolo García-Carpintero, for giving me the opportunity to do the Ph.D. at Logos. I am very grateful for his comments and advice, for being a reference in philosophical matters, and for combining this talent with kindness and personal closeness. Thanks also to my other supervisor, Josep Macià. Josep was my professor during my bachelor's and master's degrees. In his classes, the dryness of analytical thinking was always wrapped with passion and enthusiasm. I learned a lot from him, and he always made me feel heard and valid in my objections and comments. Had it not been for that early push, I might not have had the confidence to pursue this vocation. Moltes gràcies, Josep! Josep and Manolo, together with Marta Campdelacreu, were my gateway to analytic philosophy. My thinking about the matters discussed in this dissertation has been profoundly influenced by what they taught me.

I have been immensely fortunate to do the thesis at Logos. Not only because of the philosophical talent of its members and all the activities I have been able to enjoy, but also because of the character of its members. I am grateful for their feedback on this work, for all I have learned from them, and for their irreplaceable company. Special thanks go to my soul colleagues, Aaron Álvarez, Alfonso García, Carlos Benito, and Diana Couto, for their comments on this dissertation and for the beautiful time we spent together. One of the greatest academic joys in the future would be to share a department with you again. Thanks also to the rest of my colleagues, seniors and students, whom I admire. Thanks to David James Lobina, who supervised the thesis during the first years of the Ph.D., for his advice and dedication. I also want to thank the women at Logos–Josefa Toribio, Esa

Díaz-León, Genoveva Martí and Marta Campdelacreu, among others–, for being a referent from the start of my philosophical career.

The Centre for Philosophy of Memory has been my other school during the last years of the Ph.D. Special thanks to its director, Kourken Michaelian. For his advice, for reading and commenting on parts of this dissertation, and for inviting me to be an affiliated member of the CPM. I am also thankful to each of the CPM's members for their willingness to share their knowledge and comments on the contents of two chapters of this dissertation.

During my Ph.D., I have had the chance to spend research periods abroad. I want to thank Steven Sloman, for hosting me at Brown and for all our conversations on imagination, mental imagery, and other topics while walking around the Campus. I also want to thank François Recanati for hosting me at the Institut Jean-Nicod and for our conversations on imagination. Thanks also to Jordi Fernández for his advice and support along the way.

Many thanks to the people of the faculty who, with great kindness, has made it possible for the gear around each procedure to work: Luisa, Helena Gaset, Paco Murcia, and Imma Murcia.

En el terreno de lo personal, quiero dar las gracias a muchos amigos que me han ayudado a dejar de lado lo abstracto y a sobrellevar algunas dificultades a lo largo de estos años. A Laura Lores, por animarme a empezar el doctorado. A Javier Ramos, por hacerme reír y ayudarme a relativizar cuando lo he necesitado. A Giovanna Volpe, por el jolgorio y la lealtad. A mi amigo y colega de facultad Juan Evaristo Valls, por los paseos, el cariño y la sensibilidad. A Pedro Martins, por muchas conversaciones sobre la tesis y por su inestimable compañía. A Federico Viglione, por su contribución y apoyo en estados iniciales de la tesis. A Brian y Natasha, for making Providence a warm and charming place.

Y, sobre todo, muchas gracias a mi familia: a mis padres y mis hermanos, y a mi pareja. Cualquier logro, por pequeño o grande que sea, les corresponde. A mis padres: muchas gracias por la confianza inculcada, por vuestra generosidad y por el apoyo incondicional en las decisiones que he tomado. Gracias por haberme permitido ser esquiva a veces respecto a estas decisiones, y por confiar en que las cosas saldrían bien. Muchas gracias a mis hermanos; Pablo, Alba, Clara, por vuestra influencia, amor y compañía. Muchas gracias también, Clara, por haberme ayudado en una tarea que me es tan ajena como maquetar la tesis. Dedico esta tesis también a mis abuelos, a quienes me hubiera gustado abrazar ya doctorada. Por último y más importante, muchas gracias a mi pareja, Marc. Gracias por leer la tesis y por tus inestimables comentarios. Gracias también por haber compartido conmigo referencias que me podían ser de utilidad y por las muchas conversaciones sobre temas relacionados, que tan feliz me hacen. Y aunque detestes que se te agradezcan este tipo de cosas, muchas gracias también por tu apoyo incondicional en lo personal, por tu lealtad y soporte en momentos difíciles, y por enseñarme a anticipar que las cosas salgan bien. No conozco una suerte mayor que la de haber escrito esta tesis a tu lado.

***

This work was possible thanks to the financial support of an FPI grant from the Spanish Ministry of Education in the context of the project "The Significance of First-Personal Attitudes (FFI2016-80588-R)", and a grant from the Leverhulme International Network Project "What's so special about First-Person Thought?".

Contents

Part II: Boundaries of imagination

Introduction

Our capacity to entertain different worlds beyond the here and now encompasses a rich set of mental processes. When bored, we daydream about doing something else, somewhere else. When planning, we visualize the multiple steps needed to achieve the desired goal. Some believe imagination takes part in remembering or even suffering from delusion. What are the consequences and limits of imagination? Philosophers have attempted to delineate the architecture, role, and legitimacy of imagination by investigating its relationship with a wide array of mental processes: the formation of beliefs (epistemology), action in pretense (philosophy of action), delusions (philosophy of psychiatry), and memory (philosophy of mind). In this dissertation, I review the empirical and theoretical research in each of these fields, provide criticisms to the prominent theories, and, most importantly, propose new accounts of the problems at hand.

The thesis is set up as a collection of four self-standing essays, which can be read independently. Throughout these essays, I critically assess and take part in the controversies surrounding the limits and scope of imagination. The thesis is divided into two parts: *leaks of imagination* and *boundaries of imagination*. The first one, *leaks of imagination,* is devoted to the epistemological and behavioral consequences of engaging in imagination. The second one, *boundaries of imagination*, delineates imagination by raising concerns regarding it as an explanation of delusions and the nature of mental processes such as episodic memory.

Leaks of imagination

Epistemologists have questioned the legitimacy of imagination to justify beliefs. The approach has been mainly normative: should we form beliefs based on imaginings? In contrast, relatively little attention has been devoted to the descriptive question of whether experiential imagination changes beliefs. To fill this gap, **Chapter 1** explores the philosophical implications of findings on the effects of imagination. This chapter reviews clinical and psychological evidence showing that experiential imagination can influence emotional responses, attitudes, and behavior to a similar degree as perception. A plethora of successful clinical interventions (e.g., imaginal exposure in the treatment of phobias or systematic desensitization) make use of mental imagery to modulate the emotional and behavioral responses of the patient outside the therapeutic setting. Other experimental paradigms (e.g., the imagined contact paradigm) show that imagining a positive interaction or physical contact with an outgroup member reduces prejudice and intragroup bias. My goal is to argue that imagining is not an innocuous epistemic enterprise. Within the chapter, I give an account as to why, even when imaginings are correctly monitored at the personal level, there is learning in engaging with imagination. I claim that the nature of this learning is associative and happens by default when certain circumstances in quasi-sensory imaginings are met. I claim that experiential imaginings have, by default, *implicit assertoric force* and put forth a theory—the *Prima Facie* View—as a unified explanation for the empirical findings reviewed. The *Prima Facie* View concerns the architecture and functional role of experiential imagination. According to it, mental images and percepts are indistinguishable in operations involving associative and affective systems. By the end of the chapter, I address alternative strategies that could also account for the empirical evidence reviewed—such as a Spinozian model of belief formation or Gendler's notion of *alief*—and potential objections to the *Prima Facie* View.

Another domain that has been associated with imagination is the philosophy of action, particularly in research on pretense. Philosophers assign imagination a guidance role in motivating pretense behaviors. **Chapter 2** concerns the role of imagination in motivating actions that some take as pretense: intrinsic symbolic actions. These are actions involving an object that stands for an absent one and are seemingly carried out as an end in itself. Research on symbolic actions has almost exclusively been the patrimony of continental dissertations. In this chapter, I give an account of them departing from analytical literature on emotional actions. Symbolic actions entered the analytic debate as instances of emotional actions (actions done in the grip of emotion). After characterizing the phenomenon and sketching the desideratum for a theory of symbolic action, I proceed to show that neither a Humean explanation in terms of a belief-desire pair nor the mere appeal to emotions render these actions intelligible. I then evaluate Goldie's (2000) account. According to Goldie, symbolic actions are episodes of active imagination in which the subject imagines himself satisfying a desire caused by an emotion. After raising criticisms of this account, I formulate an original account of symbolic actions in which they are distinguishable from pretense. In this account, I combine the appeal to the imagination with Scarantino and Nielsen's (2016) appeal to redirected actions in the animal realm. By my account, symbolic actions are symbolically displaced active imaginings that allow for a symbolic satisfaction of a thwarted goal.

Boundaries of imagination

Chapters three and four aim to delimit the appeal to the imagination in two domains: psychopathology and episodic memory.

The non-responsiveness to evidence, circumscription, and behavioral inertness of delusions has motivated a philosophical debate on the status of delusions. The focus has been on whether these should be regarded as beliefs, as they are in psychiatry and psychology. Some authors have criticized the doxastic conception of delusion, leasing to alternative accounts. The functional profile of imaginings and more specifically, their circumscription and lack of action guidance, have been exploited in accounts of delusions. This is the case of Currie and colleagues' non-doxastic metacognitive account, according to which delusions are imaginings misidentified as beliefs by the subject suffering from delusions (Currie, 2000; Currie & Jureidini, 2001; Currie & Ravenscroft, 2002). In **Chapter 3**, I take a closer look at an account that takes monothematic delusions to be imaginings, and question Currie and colleagues' positive thesis based on new criticisms. More specifically, I show that Curie and colleagues' (2000, 2001, 2002) metacognitive account is not well equipped to explain delusional incorrigibility. Then, I explain delusions within a fragmented model of belief. In doing so, I raise criticisms of Davies and Egan's (2013) doxastic and Bayesian accounts of delusions in their fragmented system. I argue that Davies and Egan's commitment to Bayesian laws of belief formation and revision hinders their ability to explain delusions qua beliefs. This criticism is followed by a positive proposal to model delusions in a fragmented, albeit non-Bayesian, system of belief (Mandelbaum, 2019; Mandelbaum & Bendaña, 2020), which allows for the influence of motivational factors in belief acquisition and updating.

Lastly, imagination as a mental process has been at the center of debates on the ontology of mental processes. More specifically, recent theories under the umbrella of *Continuism* claim that the difference between imagination and memory is a matter of degree, not of kind. According to Michaelian's Simulation Theory, remembering an episode is simulating it in imagination, reducing memory to the act of imagining. **Chapter 4** examines Simulation Theory's ability to explain our capacity to distinguish episodic memory from free imagination. Simulation Theory suggests that we can reliably do so because of the distinctive phenomenology episodic memory comes with (i.e., a *feeling of remembering*), which other episodic imaginings lack. In this chapter, I raise two objections to the feeling of remembering as it is portrayed in the Theory. I then provide an exhaustive exploration of the theory's ability to ground the mechanism underlying this feeling. I conclude that Simulation Theory cannot simultaneously defend the simulational character of episodic memory and ground our ability to discriminate between memories and imaginings.

Part I
Leaks of imagination

Chapter 1

The *Prima Facie* View of experiential imagination

A mind that is stretched by a new idea or sensation
never shrinks back to its former dimensions.
Oliver Wendell Holmes, *Autocrat of the Breakfast Table* (1858)

Abstract. Perception is said to have assertoric force: It inclines the perceiver to believe its content. In contrast, experiential imagination—perception-like imaginings from a first-person perspective—is commonly taken to be non-assertoric: Imagining winning a piano contest does not incline the imaginer to believe that she has won a piano contest. However, plenty of evidence from clinical and experimental psychology shows that imagination can influence attitudes and behavior to a degree similar to perceptual experience. The main goal of this chapter is to argue that imagining is not an innocuous epistemic enterprise. I propose that experiential imaginings have by default *implicit assertoric force* and put forth a theory—the *Prima Facie* View—as a unified explanation for the empirical findings reviewed. According to the *Prima Facie* View, mental images and percepts are indistinguishable in operations involving associative and affective systems. I address alternative strategies that could also account for the empirical evidence reviewed—such

as a Spinozian model of belief formation or Gendler's notion of *alief*—and potential objections to the *Prima Facie* View.

1.1. Imagining experiences: World sensitivity and assertoric force

At the beginning of *Mimesis as Make-Believe* (1990), Walton introduces the following case:

> "Fred finds himself, in an idle moment, alone with his thoughts. Feeling unsuccessful and unappreciated, he embarks on a daydream in which he is rich and famous. He calls up images of applauding constituents, visiting dignitaries, a huge mansion, doting women, and fancy cars. But alas, reality eventually reasserts itself and Fred gets back to selling shoes. (…) Before proceeding we should note the independence of imagining from truth and belief. Much of what Fred imagines is false and is known by him to be false". (Walton, 1990, p. 13).

In this example, Fred uses his imagination to distance himself from reality. It is well known that this kind of imaginings can elicit emotions similar to those of actually experiencing the episode. While imagining, Fred might temporarily feel joy and hope. But, beyond these immediate emotional effects, these imaginative exercises in healthy subjects are usually regarded as epistemically innocuous from a diachronic perspective. As Walton points out, Fred knows that his imagined episode is false. Because of this, the content of his daydream will not play an evidentiary role, nor will it modify his conception of the world. The aim of this chapter is to cast doubt on this common intuition. Based on empirical evidence, I will suggest that we need to acknowledge that, by default, imagining experiences influences—to a greater or lesser extent—our implicit attitudes about the world, which can influence our beliefs and affect our behavior. This occurs by creating associations that are extended from imaginings to real stimuli. According to this, after his fantasy lapse, Fred returns to work having—at least slightly—

modified his attitudes about the world. For instance, he might now take the imagined episode to be more probable (Carroll, 1978), idealize the goals represented and consider them easier to achieve (Kappes, Oettingen, & Mayer, 2012), or slightly actualize his self-concept in a positive manner as a result of being associated with success (Kappes & Morewegde, 2016).

Following Walton's observation in the initial fragment, some claim that perception inclines the perceiver to believe its content—it has assertoric force—while imagination does not—it has non-assertoric force (Chasid & Weskler, 2020). Exceptions to the non-assertoric force of imagination are attributed to reality monitoring errors. Two kinds are commonly identified: Hallucinations—an internally generated experience is evaluated as real, leading to similar doxastic consequences as perception (Dijkstra, Kok, & Fleming, 2022)[1]—and imagination inflation—a phenomenon in which the subject mistakes an episode or action that she imagined for one that actually occurred (Garry et al., 1996; Goff, 1998; Loftus, 2003). These are suboptimal circumstances in which imagination will have assertoric force and lead to the formation of beliefs. Besides these error cases, imagination can also influence belief in a straightforward way when the subject decides to use it for an epistemic purpose. For instance, we can use sensory imagination to determine whether a river is crossable (Harahan, 2021): If in the imagining it appears crossable, we may form the belief that it is.

But, apart from the reality monitoring errors and epistemic uses of imagination already acknowledged in the philosophical literature, the norm is that engaging in experiential imagination does not necessarily have an influence in the epistemic status of the subject. In this chapter, I want to support the idea that, to some extent, experiential imagination influences the epistemic status of the subject in more insidious ways than one would assume.

[1] I refer to hallucinations without insight. I leave here opened the possibility of perceptual hallucinations which are *experienced* as real while *knowing* they are not real. Hallucinations that take place after drug consumption exemplify this case.

A large body of empirical evidence shows that, even in the absence of reality monitoring errors, imagination can have similar consequences to its perceptual counterpart. Among the evidence presented, clinical uses of imagination have been in practice for a long time, but their philosophical implications have gone unnoticed. Such is the case of imaginal exposure, a well-established cognitive-behavioral treatment used with phobic patients as a first step or "warm-up" toward a strongly feared in vivo exposure (Anthony & Swinson, 2000). In this technique, although the approach to the fear-inducing stimulus is merely imagined, the training affects the extinction of the conditioning to a similar degree as in vivo exposure (Hackmann, Bennett-Levy, & Holmes, 2011). This behavioral effect alone—perhaps the most astonishing and consequential—is hard to reconcile with the classical view of imagination. And yet, this is only one piece of evidence from a larger body in psychology that undoubtedly demonstrates that, without the presence of any monitoring errors, imagination can cause psychological and behavioral reactions akin to those of the perceptual experience (Morewedge et al., 2010; Shidlovski et al., 2014; Szpunar & Schacter, 2013).

After reviewing this work, I will argue that experiential imaginings have *implicit assertoric force*, which constitutes a tenet of the view of imagination—the *Prima Facie View*—introduced in the chapter. To summarize, the *Prima Facie* View defends that in operations at the subpersonal level, mental images involved in experiential imaginings are processed *at face value* by default. Namely, these representations are processed by their mere appearances (*prima facie*), regardless of background information about their source or representational nature.

But, before we get there, let me first clarify the notion of experiential imagination with which we will be working. The notion of imagination that is operative throughout the chapter has also been named perceptual or quasi-sensory imagination (Nanay, 2015). This refers to imaginings within which we have a particular experience by means of mental imagery (i.e., representations and the accompanying experience of sensory information without a direct external stimulus; Pearson et al., 2015).[2]

In what follows, I compare two views on the architecture and functional role of experiential imagination. I take the first one to be preponderant in the literature. Since, under normal circumstances, it does not attribute default epistemic consequences to imagination, I will call it the Innocuous View. The second one is the view I will sketch in this chapter. In Section 2, I present empirical evidence on the attitudinal and behavioral effects of imagination. This section constitutes the bulk of the chapter and motivates the proposed functional theory of imagination. Ultimately, I argue that the *Prima Facie* View can unify and better explain the heterogeneous phenomena presented. Therefore, it is a plausible candidate for the cognitive architecture and functional role of imagination that demands further investigation. Section 4 is devoted to formulating the theory in some detail. Sections 5 and 6 present alternative explanations of the phenomena reviewed—Spinozian Theories of belief formation and Gendler's notion of *alief*—and possible objections to the *Prima Facie* View.

1.2. The Innocuous View vs. the *Prima Facie* View

It is almost a platitude that healthy subjects can fantasize and indulge themselves in daydreaming without giving evidentiary value to the contents represented in imagination. That is, they can entertain experiential imaginings without this straightforwardly influencing their attitudes about how the world actually is. To take just an example, Wittgenstein writes:

[2] Imagination is frequently characterized as having a perception-like and a propositional variant (Schellenberg 2014: 499). An example of the first is imagining *submerging* in the river Ouse—visualizing the river, the sky, and so forth. The second kind of imagination concerns imagining a state of affairs being actual, such as that Caesar's troops crossed the River Ouse during the Gallic War. Here, I am only concerned with the first variant.

> "It is just because forming images is a voluntary activity that it does not instruct us about the external world." (Wittgenstein, 1967, p. 621)

As Wittgenstein's words point out, it is commonly advocated that since imagination—unlike perception—is not sensitive to how the world is—that is, it does not track changes in it—, it is not a suitable tool for acquiring new evidence about the world (Badura & Kind, 2021). Unlike Wittgenstein, others consider that experiential imagination can have legitimate epistemic uses, at least in some cases. For instance, it has been claimed that we can use perception-like imagination to determine whether a river is crossable (Harahan, 2021). I will not address the normative question here of whether imagination can be legitimately used for an epistemic enterprise. I will limit my inquiry to the descriptive domain, that is, to the question of whether imagination influences *de facto*, to some extent, our attitudes about how the world is.[3]

I will refer to the perspective that imagination does not necessarily influence our attitudes about the world and with it our epistemic status, the Innocuous View. If supporters of the Innocuous theory had an anthem, it would be Dreaming by Blondie, which chorus says, "Dreaming is free." Apart from reality monitoring errors—which cause imagination to influence our beliefs and actions—and epistemic uses—imagination is seen as a playground where one can entertain oneself endlessly without epistemic consequences on one's return to reality. It is important to note that the Innocuous View does not address the normative claim that we *should not* treat imaginings as observed evidence, but rather the descriptive claim that that we *do not* normally treat them as such. In sum, the crucial aspect of the Innocuous Theory is that it takes the functional profile of imagination to match its normative epistemic profile under normal circumstances. Going

[3] Moreover, the phenomena I will present here are not epistemic uses of imagination. In them, the imaginer does not attempt to discover what the world is like through imagination, and it would be odd to defend that such imaginings can have an evidentiary role in determining how the world is.

back to the previous example in which Fred imagined being famous, he should not, and, under normal circumstances, he *will not* give it evidentiary value nor change its mind about how the world is based on such imagining because their content is under his control and not constrained by how the world is.

I contend that the Innocuous View is implicitly the predominant view in the literature on experiential imagination. Namely, imagination affecting attitudes and behavior is taken to be an exception to its normal functional profile. There are accounts of specific cases in which imagination can lead to actions—mainly concerning the context of pretense—precisely because it taken to be exceptional (Schellenberg, 2014). A recent line of inquiry has explored the relationship between our imaginative capacities, beliefs, and emotions (Nichols, 2004; Stock, 2017; Weinberg & Meskin, 2006). This line has been inspired mainly by our engagement with fiction and active pretense (Schellenberg 2013; Gendler, 2010; Langland-Hassan, 2012). Although these authors have focused on imagination in its propositional variant, they have an implicit Innocuous View of experiential imagination in the epistemic domain.

I will refer to exceptions to the preponderant Innocuous View as *Overshooting accounts*. These accounts, in the attempt to explain similarities between imagination and belief (such as, for instance, the emotions elicited by imagining) have either posited new mental states—Gendler's notion of *alief* (2010)—, claimed that imagination and belief are in the same code (Nichols, 2014), or argued that the mere activation of a truth-apt proposition leads to immediately believing it—Spinozian models of belief formation (Gilbert, 1991; Mandelbaum, 2014, p. 55). Although they accommodate for the similarities between belief and imagination, those accounts fail to explain crucial discontinuities between believing and imagining.[4] Furthermore, by

[4] I cannot evaluate all such alternative theories here for space reasons. In section five, however, the Spinozian Theory and Gendler's notion of *alief* are examined as alternative accounts of empirical evidence reviewed.

most of these accounts, mental images do not have a role in mediating the effects of imagination, a claim that the view here presented departs from.

I claim a more parsimonious view can be achieved by appealing to the link between imagination and the associative system. What I call the *Prima Facie* view consists of the following claims:

1) When imagining an experience, a subsystem (the *Prima Facie* system) processes the mental imagery *at face value*, ignoring the fact that the source of this representation is not world-sensitive. This system is cognitively isolated from evaluations on the source of the representation (perception, virtual reality, a film, imagination, et cetera), and by default gives perceptual force to the contents entertained.

2) Imagination has, by default, *implicit assertoric force*: On many subpersonal operations, its contents are processed ignoring source information and therefore integrated as observations with evidentiary value (similar to perceptual experiences). Because of this, the parade of sensory contents of imaginings—mental images, affective responses—not only triggers associations but can also create them, similarly to actual experiences. These associations can generalize from imagery to real stimuli resulting in implicit attitudes, affecting behavior, and sometimes influencing beliefs.

3) These effects are independent of reality monitoring errors, and they take place even if, at the personal level, we correctly identify imaginings as such when having them and when remembering them.

4) The *Prima Facie* system encompasses at least the associative system.[5] The functional profile of imagination regarding this system explains why entertaining experiential imaginings can sometimes have

[5] It is beyond the scope of this chapter to determine the whole range of systems that might process the contents of experiential imagination *prima facie*, such as the systems responsible for many physiological responses that can take place when imagining (e.g., systolic blood pressure—Kappes and Oettingen 2011, or skin conductance—Mueller et al., 2019).

> attitudinal and behavioral consequences that resemble those of experience.

The described functional role would obey the architecture of the imagination, and the effects I will review in the next section, although not always desirable, would reflect the operations of a well-functioning system. The phenomena described in the next section constitute a sample of what I name *Prima Facie* effects. As said, I take them to be due to the contents of imaginings being processed *at face value* in subpersonal operations. In other words, they are treated as if they had originated in the actual co-occurring situation and thus were the product of perceptual experiences and constitute evidence about the world.

1.3. Phenomena to be explained

1.3.1. Imaginal exposure as a treatment for phobia

A phobia is an unrealistic fear of a situation, person, or object. Phobias are usually explained by conditioning models of fear acquisition (Field, 2006 Watson & Rayner, 1920). In conditioning terminology, phobias originate when a previously neutral stimulus is paired with an aversive unconditioned stimulus.[6] For instance, a patient may have a phobia of dogs because she was once bitten by one. The bite is the unconditioned stimulus, and the fear response is the unconditioned response. Through its association with the unconditioned stimulus, the neutral stimulus (dogs) becomes conditioned (conditioned stimulus). As a result, the conditioned stimulus then elicits the same fear response as the unconditioned stimulus. The fear of the conditioned stimulus becomes the conditioned response. Individuals who have a phobia tend to avoid or run from situations where the phobic, fear-

[6] The unconditioned stimulus receives this name because it evokes an anxiety or fear response (the unconditioned response) without the need for any learning or conditioning. An example of an unconditioned stimulus would be a dog bite, which tends to cause an immediate fear response.

inducing stimulus is present. This flight response, in turn, is reinforced by the subsequent reduction in fear and anxiety (i.e., negative reinforcement), which contributes to the persistence of the phobia. Avoiding the phobic stimulus prevents the subject from learning from staying in the situation—namely, learning that her feared outcomes will likely not come true or that, if they do, they are not as terrible as she imagines.

Learning that the phobic stimuli are not threatening is the purpose of exposure therapy. This behavioral technique involves a set of stages in which the patient approaches the fear-inducing stimulus gradually. If the fear-inducing stimulus is presented alone (namely, without predicted negative outcome), the conditioned response's strength will decline over successive trials until the stimulus no longer elicits fear. Exposure therapy leads to the successful extinction of fear and has become the most popular treatment for specific phobias (Eaton et al., 2018).

The point I want to emphasize here concerns the efficacy of *imaginal* exposure therapy (Hackmann et al., 2011; Rentz et al., 2003). This modality of therapy differs from *in vivo* exposure in that the exposure is merely imagined. In therapy, patients are asked to visualize, in detail and as vividly as possible, a confrontation with the fear-inducing stimulus for a certain amount of time—ideally until fear and anxiety begin to subside.[7] Following our previous example, a patient would imagine interactions with dogs where she is not bitten. The mere imaginal confrontation with the feared stimulus is effective in inducing a fear response (Grayson, 1982) and contributing to fear extinction—the lessening of the conditioned fear response—similarly to *in vivo* exposure (Choy et al., 2007; Wolitzky-Taylor et al., 2008).[8]

[7] Imaginal exposure offers many advantages over *in vivo* exposure. It is more convenient (it can be easily conducted in a therapist's office) and more flexible (imaginal techniques can be adapted to fit the idiosyncratic situations that evoke the patient's fear). Imaginal procedures also allow a gradual confrontation with fearful situations that can prepare the patient for *in vivo* exposure.

[8] Other imagery techniques with phobias have also proven effective. For example, systematic desensitization, in which the patient is trained to relax their voluntary muscles during the imaginal confrontation with the feared stimulus (Rachman, 1967). Imaginal

These findings appear mysterious from the perspective of the Innocuous View. Crucially, there would not have been any actual, new evidence of the harmlessness of dogs during imaginal exposure: The patient has just been exposed to an imagined dog, and the imagining was under her control. More importantly, patients correctly monitor such episodes as imagined during and after—when recalling—the imagined episodes. And yet, merely imagining an interaction with the phobic stimulus influences what the patient will anticipate and how she will behave in the future when presented with the phobic stimulus. Experiential imagining has a similar effect to perception—or *in vivo* exposure—in the extinction of the fear response. The fact that imagining a positive interaction with the phobic stimulus is as effective as interacting with the physical stimulus to extinguish conditioning has already been pointed out in the past (Dadds et al., 1997; Mertens et al., 2020), although this has gone unnoticed by the philosophical literature on imagination.

The *Prima Facie* View can accommodate and, in fact, predicts these findings. Acknowledging that the contents of imaginings are inputted to the associative system without information about their source accounts for the efficacy of imaginal exposure. First, it accounts for the fact that the mere image of a dog causes a fear response similar to the perception of the dog (even if the subject acknowledges that she merely imagines it and that she is safe). The associations that the perception of a dog would trigger are also evoked by the mental image of the dog. Second, it explains why imagining a harmless dog can break its association with the conditioned stimulus (the bite). According to the *Prima Facie* theory, in the imaginative exercise the mental image of the dog is paired with the reduction of the fear. Because of the similarity between the mental image and the precept of a dog, the associations resulting from the imaginal exercise are generalized from mental

techniques have also been used successfully in patients with PTSD (Arntz et al., 2007; Bryant et al., 2003; Minen & Foa, 2006) and obsessive-compulsive disorder (Abramowitz et al., 1996; Foa et al., 1980). Additionally, in recent years imaginal exposure through virtual reality has also been proven highly effective (Botella et al., 2017).

imagery to reality. This explains the reduction of fear and the behavioral changes in the presence of the real fear-inducing stimulus after therapy.

1.3.2. Fantasies of self-achievement and motivation

The Hopes so juicy ripening-
You almost bathed your Tongue-
Emily Dickinson

Another interesting phenomenon concerns the effects of imagination on motivation and achievement. In "The Undoing Project," Michael Lewis reports a rule Daniel Kahneman established for himself in his childhood:

> "As a child during the war, he'd cultivated an active fantasy life. He would play out elaborate scenes with himself at the center of them. He imagined himself single-handedly winning the war and ending it, for example. But because he was Danny, he made a rule about his fantasy life: *He never fantasized about something that might happen.* He established this private rule for his imagination once he realized that, after he had fantasized about something that might actually happen, he lost his drive to make it happen. His fantasies were so vivid that 'it was as if you actually had it,' and if you actually had it, why would you bother to work hard to get it?" (2016, p. 443; emphasis mine)

At first glance, Kahneman's rule may seem counterintuitive. Many of us would introspectively find that optimism about the future would boost our motivation to achieve a goal. In this line, research shows that thinking positively about the future increases motivation and performance (Bandura, 1997). However, thinking positively about the future is different from indulging in self-achievement fantasies. Empirical studies on motivation differentiate between two ways of thinking about the future: positive expectations—i.e., judging the desired future as likely—and positive fantasies—i.e., experiencing one's fantasies about a desired future positively (Kappes & Morewedge, 2016). Positive fantasies—the kind that Kahneman avoided—predict low success and effort in several domains (Kappes et al., 2012). Oettingen and Wadden (1991) investigated the impact of expectations and fantasies in a one-year behavioral weight reduction program. They found that these variables, measured pretreatment, predicted weight change in opposite directions. Optimistic expectations of reaching one's goal combined with weight-related negative fantasies favored weight loss. On the contrary, subjects who displayed pessimistic expectations combined with positive fantasies had the poorest treatment outcome. Along the same lines, Oettinguen and Mayer (2002) found that positive expectations about earning a high grade predicted high effort and successful performance, but positive fantasies of achievement predicted low effort and performance. Similarly, positive future fantasies have been shown to predict lower grades at the end of the academic program (even controlling for academic competence, expectations of achievement, and self-discipline; Kappes, Oettingen, & Mayer, 2012). Even when positive fantasies are induced, they still result in a lower energy investment than more pessimistic or neutral fantasies (Kappes & Oettingen, 2011). Note that these effects occur even when subjects correctly identify imaginings as such during and after the procedure. That is, imagining winning a piano contest—even if we visualize it in detail and vividly—will not lead us to believe that we have won a piano contest, nor will it inflate the confidence we have that the event has occurred (i.e., imagination inflation). In sum, to indulge oneself in images of a bright future, even when correctly monitoring these fantasies, might have effects similar to experiencing actual achievement, such as decreasing motivation.

In evaluating this line of research, Kappes and Morewedge (2016) consider two mechanisms to account for these effects. First, they claim that imagined achievement may sometimes serve as a substitute for real achievement. Since the simulation of success generates similar feelings, the need to devote effort to producing real success decreases. In their words, "mental simulations make people feel, to some extent, like that event has actually happened. (…) Just as people give themselves credit for their good intentions (Kruger & Gilovich, 2004), mentally simulating success may allow people to feel successful without effortfully pursuing their goals" (2016, p. 413). Second, they attribute the decrease in achievement to the fact that idealizing the future omits obstacles and challenges, which hampers planning about how to proceed in order to achieve the desired outcome in real life.

The *Prima Facie* View accounts for both effects. Concerning the positive affect, this would be elicited by the associative system which reacts to the parade of imagined contents in a similar way to perceptual experience—since it processes its contents *prima facie*. Therefore, the prompted emotions are similar to the ones of perceptually experiencing real achievement.[9] Beyond the affective response, fantasies of self-achievement might influence our self-concept by increasing the association between oneself and success. Fantasies that depict an idealized version of future events might end up influencing our motivation by changing our self-concept. If we intend to achieve certain goals to improve our self-concept or feel better about ourselves and we can achieve these same effects by imagining, this last option could be preferred and therefore lower our motivation to *actually* achieve such things.

Concerning the idealization of the future, it is essential to take notice that subjects seem to somehow be giving evidentiary value to the imagining—for example, after imagining it can seem easier to achieve the imagined outcome. The *Prima Facie* View can account for such evidence by appealing to the following: By easily invoking a visualization of our success, we might associate the ease in generating the imagery with the output represented in it. This is in line with The Simulation Heuristic (Kahneman

[9] Differences in the degree of the emotion elicited by the imagining and the real experience may be accounted by a posterior mechanism that integrates information about the source of the representation.

& Tversky, 1981), according to which "the ease with which the simulation of a system reaches a particular state is eventually used to judge the propensity of the real system to produce that state." Although the Simulation Heuristic was formulated to account for the perceived plausibility of counterfactual events, this evidence suggests that the heuristic could also play a role in evaluating the plausibility of future events. It is possible that in imagining a particular desired outcome, we associate its representation with the ease with which we can imagine it. Furthermore, repeatedly imagining such an event might increase the fluency of future imaginings of the event, which might then influence our judgement on how easily we can achieve the outcomes represented. Because the system generalizes the associations of mental images to real stimuli represented by them, we could end up associating the desired outcome with "ease of achievement," which might decrease our motivation by not being challenging anymore.

1.3.3. Imagined Contact and prejudice reduction

According to the Contact Hypothesis (Allport, 1954), interaction between members of opposing groups leads to more positive outgroup attitudes and lessens hostility—under the right circumstances. A recent meta-analysis shows a robust effect of contact on prejudice (Pettigrew & Troop, 2006). In the last decade, there has been a turn towards the Imaginal Contact Hypothesis (Crisp & Turner, 2009). According to this hypothesis, merely imagined contact with outgroup members improves intergroup attitudes. Due to segregation or intergroup conflict, members of different groups do not tend to interact. This imaginal technique is used as either a preliminary approach or an alternative intervention—when real contact is not possible—in order to increase contact with members of the outgroup and improve relations. In the Imaginal Contact paradigm, participants in the experimental condition are asked to imagine a positive interaction with an outgroup member. Here is an example of what participants assigned to the *imagined contact condition* are asked to imagine:

> "One day you find yourself on a busy train. You get a seat and start reading the novel you brought with you to pass the time. At the next stop, an older Black man boards the train and sits down next to you. After a few minutes, the man looks at what you are reading and comments that it is one of his favorite books. This begins a discussion in which you share your thoughts on the book and what you both enjoyed about it. The conversation meanders, and by the time you get off the train, 30 minutes later, you have discussed a whole range of topics, from the stresses of having to commute to work every day, to what neighborhood you live in, to what your children's favorite subjects are at school." (Crisp & Turner, 2009, p. 231)

A large number of experiments demonstrate that simulating social contact with an outgroup member is sufficient to improve intergroup attitudes (for a review, see Miles & Crisp, 2014). Turner et al. (2007) found that young people who imagined a positive encounter with an older person showed lower levels of intergroup bias than participants who imagined an outdoor scene (Experiment 1) and participants who simply thought about older people (Experiment 2)[10]. Along the same lines, heterosexual men who imagined a positive interaction with a gay man in which they learned something subsequently had a more positive attitude toward gay people in general. In Turner and Crisp (2010), non-Muslim participants who imagined talking to a Muslim stranger subsequently showed more positive implicit attitudes towards Muslims (as measured by a Muslim/non-Muslim version of the IAT), compared to the control condition (imagining a hiking trip). Imaginal contact has also been used to change the attitudes of people high in right-wing authoritarianism (RWA; Asbrock et al., 2013). After imagined contact, participants high in RWA showed fewer negative emotions toward Turkish people (Study 1) and more willingness to engage in future contact with Romani people (Study 2). The effect has also been shown in children.

[10] Measured by their reported preference for being paired in a future study with another young person or an elderly person (on a 9-point scale).

Compared to the control group, children who simulated contact with a child with a disability (Cameron et al., 2011) subsequently showed reduced intergroup bias in their general attitude and ratings of warmth and competence. In children between 5 and 6 years, imagined contact also led to more positive intended friendship behavior towards children with disabilities.[11]

The literature so far does not report reality monitoring errors in these imaginal exercises: Subjects correctly monitor them. The *Prima Facie* View can account for these effects by appealing again to how imaginings influence the associative and affective systems. In this case, engaging in positive experiential imaginings with an outgroup member will improve attitudes towards the outgroup through the same associative process as evaluative conditioning.[12] By imagining a positive interaction with a member of the outgroup, the valence of the outgroup becomes more positive and reduces intergroup bias paralleling the effects of face-to-face contact. This is because in the functional profile of imagination, contents are processed indistinguishably from precepts in subpersonal operations. Therefore, they have similar effects to observations coming from experience. Because of this, they can generate associations—for instance, between the mental image of an outgroup member and a positive emotion. Because of the source-indifference of the associative system, these associations between the mental image and the emotion are automatically generalized to members of that

[11] The imaginal contact hypothesis has also been tested by simulating physical contact with an outgroup member. In Shamloo et al. (2018), participants were divided into two conditions: intergroup physical condition (InterPC) and intragroup physical condition (IntraPC). Participants in the InterPC condition were asked to imagine touching the hand of an outgroup member (an African-American individual). Participants in the IntraPC condition were asked to imagine touching the hand of an ingroup member (a Caucasian individual). While doing so, they were asked to "imagine feeling at ease during this contact and imagine it to be a positive experience in which you discover unexpected things." As a result, participants in the InterPC condition showed lower intergroup bias levels than those in the IntraPC condition.

[12] Evaluative conditioning is the process by which a stimulus (the conditioned stimulus) is paired with a positive or negative unconditioned stimulus which changes the evaluation of the conditioned stimulus.

outgroup, mimicking the effect of a real positive experience with them. Recent empirical evidence on evaluative conditioning and imagery supports the interpretation offered by the *Prima Facie* View. Evaluative conditioning can occur with voluntarily formed visual imagery in place of perceptual stimuli. Lewis et al. (2013) showed that voluntary mental images become conditioned when followed by emotion-evoking stimuli (pictures with positive valence). More importantly, they showed that the conditioning generalized from the mental image to the real stimulus. After conditioning the mental image, they found that perceptual stimuli of the same content produced the associated emotional response. This is in line with the implicit assertoric force of imagination: The effects are due to the associative system (part of the *Prima Facie* System) not integrating information about the source of the representation being internal and treating it by default as a perceptual force. That is, it reacts to it *prima facie*. Because of this, it does not have one set of associations for the internally triggered mental image of X and a different one for the associations triggered by perceiving X. The associations elicited by both are the same. Given that, associations created when imagining will therefore be at play in real interactions with the imagined stimulus.

1.3.4. Imagination and probability judgements

A frequently acknowledged epistemic use of imagination is as a guide for knowing possibilities (Kung, 2010; Yablo, 1993).[13] This claim could be defended by many partisans of what I call the Innocuous View. However, this section concerns a more compromising phenomenon: the influence of imagination in probability judgments, and not its contribution to assessing possibility. In short, imagining an event influences the imaginer's estimation

[13] Hume's claim in the Treatise exemplifies the case: "This is an established maxim in metaphysics. That whatever the mind clear conceives includes the idea of possible existence, or, in other words, that nothing we imagine is absolutely impossible" (I.ii.2).

of the probability that an event will happen. Imaginings, even when correctly monitored, seem to be given evidentiary value in certain subpersonal operations, influencing probability estimations as real observable evidence would.

Experiments have demonstrated that vividly imagining the occurrence of a particular event increases its subjective probability. Before the 1976 Presidential election was held, voters assigned to imagine Carter winning predicted that he was more likely to win than voters assigned to imagine Ford winning (Carroll, 1978). Similarly, subjects asked to imagine a good football season for a team were more likely to predict a major bowl bid for such a team than subjects asked to imagine a bad season (Carroll, 1978). Imagining also influences our intention to perform a specific action. Research participants who imagined donating blood, changing their major, or taking a vacation exhibited increases in their expectations of doing so (Anderson, 1983). Imagining not only changes intentions but also behavior, as recently demonstrated: Imagining oneself voting increases the probability of voting (Libby et al., 2007). Similar effects have been found in a consumer context. Gregory et al. (1982) gave information about a cable television service to only half of the residents in a neighborhood and asked the other half to imagine themselves utilizing it. Several weeks later, the cable company requested these residents' orders for cable service. As a result of the previous intervention, 19.5% of the residents who had only heard about the product's features subscribed to the service. Surprisingly, the subscription rate was 47.4% among those that imagined enjoying the cable TV service. Similar effects have been found concerning morality: Imagining performing harmful actions makes people report a higher likelihood of performing those actions in the future (Morris, O'Connor, & Cushman, 2022).

These effects have been classically attributed to the "availability heuristic" (Sherman et al., 1985; Tversky & Kahneman, 1973). When individuals estimate the probability of an event happening, this is based on the ease of accessing or imagining relevant instances consistent with the outcome estimate. In Tversky and Kahneman's words: "Availability is an ecologically valid clue for the judgement of frequency because, in general, frequent events are easier to recall or imagine than infrequent ones" (1973, p. 209). Crucially, in the evidence reviewed, subjects do not seem to disentangle imagined events from perceived ones when judging probabilities. The previously imagined events are taken as available instances in the probability estimation, increasing fluency in recalling or imagining the estimated event. Even if, at the personal level, these imaginings are acknowledged as such, people seem to subpersonally attribute evidentiary value to the mental simulation of these events (Kappes & Morewedge, 2016).

The *Prima Facie* Theory predicts that imaginings will affect judgments of probability. In doing so, sequences of mental images create associations as perceived stimuli do: the existence of this associations increases the ease of retrieval and the availability of certain representations (mainly by strengthening the associations between sequences of images). This strengthening of associations and the attribution of evidentiary value to them even when originated in imagination is supported by recent experimental evidence. In Shidlovski and colleagues (2014), participants underwent a guided-imagination procedure in which they were asked to imagine an event (piking a specific card). [14] Subsequently, researchers measured what they called the Implicit Truth Value (ITV) of such an event. For this purpose, they used the autobiographical Implicit Association Test (Sartori, Agosta, Zogmaister, Ferrara, & Castiello, 2008). This test is frequently used to assess the truth of autobiographical events in an implicit way. It assumes that when the response to sentences related to a true autobiographical event share the response key with other true sentences, reaction time will be faster than when the response to sentences related to a true autobiographical event and

[14] A fragment of the guided-imagination task: "Imagine that there are two cards lying face down in front of you/ You pick up one of the cards/And see the 4 of diamonds/ You look at the red diamonds/ Two are placed one beside the other on the upper half of the card/ And two are on the lower half of the card/ You see the four of diamonds clearly…." (2014, p. 519)

false sentences share the same key. In the experiment, participants responded faster (greater ease) when the sentences about the imagined event shared the response key with true sentences (e.g., *I am in front of a computer*) than with false sentences (e.g., *I am climbing a mountain*). Authors conclude that imagining an event increases its *implicit truth value*, like experiencing the event does. This happens even when people acknowledge, at the personal level, that the event did not occur. Importantly for experiential imagination theories, in a different experiment, Shidlovski et al. (2014) showed that imagined representations generated from a first-person perspective—mimicking experience—had a higher implicit truth value than those generated from a third-person perspective.

1.3.5. Imagery involving other sensory modalities

Although the evidence so far has focused on visual mental imagery, evidence in other sensory modalities (e.g., auditory and somatic) supports the implicit assertoric force of imaginings. For instance, in a series of striking studies, Morewedge et al. (2010)[15] examined the effects of repeatedly imagining the consumption of food on subsequent behavior. They hypothesized that although the common intuition is that imagining eating something sensitizes oneself to it—namely, increases the appetite for such food—, mentally simulating an experience that is more analogous to prolongated consumption might engender habituation to the stimulus (that is, a decrease in one's responsiveness to the food and motivation to obtain it). In Experiment 1, participants were divided into three conditions. They were all asked to imagine performing 33 repetitive actions, one at a time. Participants in the control condition imagined inserting 33 quarters into a laundry machine. Participants in the three-repetition condition imagined inserting 30 quarters into a laundry machine and then imagined eating three M&M'S. Participants in the 30-repetition condition imagined inserting three quarters

[15] Replicated by Camerer et al. (2018).

into a laundry machine and then imagined eating 30 M&M's. After doing so, participants in all three conditions could eat *ad libitum* from a bowl containing M&M's. As a result, participants in the 30-repetition condition ate significantly fewer M&M's than participants in the three-repetition and control conditions. The authors attribute the effect to the influence of imagined consumption, concluding that habituation to food can occur by the mere imagined consumption of that food.

Another interesting phenomenon involves auditory imagery and self-talk. Research has shown that a subtle grammatical difference in our inner monologue can impact in our performance, intentions, and emotion regulation. This subtle difference refers to using first person pronouns versus other pronouns when talking to ourselves. For instance, saying to ourselves "I can keep going" versus "you can keep going." Using the second-person pronoun instead of the first-person pronoun has been shown to influence physical and cognitive performance positively. Participants in a 10 km cycling time trial performed better following second versus first-person self-talk (Hardy et al., 2019). Similarly, participants who used non-first-person pronouns and their own name during introspection performed better on the speech task (Kross et al., 2014). Using second-person self-talk in preparation for an anagram task enhanced performance and intentions to work on anagrams more than first-person self-talk (Dolcos & Albarracín, 2014). Silently talking to oneself in the third person—using one's own name—has been shown to enhance emotion regulation and facilitate self-control (Moser et al., 2017).

This phenomenon is often explained by subject increased self-distance with the use non-first personal pronouns (Hardy et al., 2019; Kross et al., 2019). However, in this metaphorical wording it is unclear what it is to distance oneself *from oneself*. Others (Moser et al., 2017) point out that third-person self-talk leads to thinking about the self as they would think about others. While I cannot elaborate here on why these explanations are not satisfactory, my goal is to show that the *Prima Facie* View provides a plausible explanation of the phenomenon by appealing to the implicit assertoric force of auditory imagery. Because the *Prima Facie* System processes auditory inputs regardless of the source—namely, at face value—, *hearing* cheers in the second person gives the impression that some person other than yourself is acclaiming you. This is because spontaneous self-talk is almost always formulated in the first person (e.g., "I can do this") and rarely in the second person, whereas the second or third person (e.g., "You can win") is associated with others talking to you. By inner second-person self-talk, subpersonal processes might react as if some other person apart from yourself believes that you can keep going. That is, the auditory imagery is processed as external judgment, which increases your confidence in being able to keep going (since, apart from yourself, someone else believes you can win). Therefore, these influences are mediated by the *prima facie* impression of someone telling you that you can do it. Someone telling himself in inner speech "You can do it" gives the *prima facie* impression that someone other than himself is telling him "You can do it."

1.4. The *Prima Facie* System

We have demonstrated by reviewing a plethora of findings from different traditions that imaginings affect attitudes and behavior *as if* they were perceptual experiences. These findings cannot be explained by the Innocuous View of imagination, since it establishes that imagination leaves no psychological footprints. The *Prima Facie* View accounts for this evidence by proposing that experiential imaginings have implicit assertoric force. That is, imagined contents are treated the same as world-sensitive precepts in certain processes, which then leads to similar effects as observations coming from experience. I will call the system that processes imaginings independently from the source of the representation the *Prima Facie*

System.[16] This system reacts to mental imagery and other contents of imaginings as if they originated in the perception of our actual, co-occurring situation. This also applies to the perception of fictional depictions (namely, not seeing a lion but seeing a depiction of a lion in a film, in Virtual Reality…), to which the *Prima Facie* system attributes perceptual force. Because of this, learned associations are triggered as a consequence of experientially imagining, with new associations being created and old associations reinforced. This then extends to the real stimulus—person, object, or situation. As a result of this process, imagination can affect behavior and beliefs (e.g., beliefs about the probability of an event happening or about how difficult a task is to achieve). In this section, I will characterize how mental images are processed regardless of source information in the *Prima Facie* System, sketch a plausible way in which the contents of experiential imaginings input the *Prima Facie* System, and explain why the reviewed evidence constitutes a unified phenomenon and specifies the mechanism underlying the *Prima Facie* effects.

1.4.1. Magritte's deference and implicit assertoric force

By saying that experiential imagination has implicit assertoric force, I mean that the parade of sensory contents of imaginings—mental images, affective responses—not only triggers but also creates associations that are treated as having evidentiary value concerning how the world is. These associations can generalize from imagery to real stimuli resulting in implicit attitudes, affecting behavior, and sometimes influencing beliefs. These effects are due to the fact that the mental image is processed *prima facie* in some subpersonal operations. The claim that the contents of imaginings are processed *prima*

[16] There are reasons to claim that this system encompasses, at least, the associative system. An open question remains regarding the relationship between the *Prima Facie* system and the associative system since they share certain characteristics (such as reacting to mental imagery *prima facie*). I take the stance that the *Prima Facie* system encompasses the associative system, but that they are independent. However, this goes beyond the scope of this chapter and will not be further addressed.

facie can be intuitively grasped in the following way: Mental images are processed *prima facie* in certain subpersonal operations because they are not accompanied by what I will call "Magritte's deference". By "Magritte's deference" I mean the following: In the famous painting The Treachery of Images (1929), an image of a pipe is shown. Below the pipe, Magritte wrote, "Ceci n'est pas une pipe", French for "This is not a pipe". When asked about the painting, Magritte answered:

> "The famous pipe. How people reproached me for it! And yet, could you stuff my pipe? No, it's just a representation, is it not? So, if I had written on my picture 'This is a pipe', I'd have been lying!"[17]

The pipe in Magritte's painting is accompanied by information on its representational nature, that is, on the fact that it is just an image and is not a pipe in and of itself, but merely represents a pipe. The pipe represented in the painting cannot be smoked, and the inscription in the painting highlights the fact that such representation is not itself a pipe. This information is precisely what mental images, when inputted to the *Prima Facie* system, lack. Namely, they do not contain information on their simulational nature, and as a result, are treated as precepts. That is, information on whether they are being caused by the perception of a pipe, internally generated—i.e., imagined—, or perceived on a screen—e.g., when seeing a movie or in Virtual Reality. That is, in subpersonal operations, they are processed without the source information marking their origin outside perception, and as a result are taken as if they had come from perception. This is what the implicit assertoric force of imaginings refers to.

Next, I sketch a plausible way in which the contents of experiential imaginings input the *Prima Facie* System, leading to the reviewed effects.[18] For this, I will use Kosslyn's model of imagery (Farah, 1984; Kosslyn, 1980)

[17] Torczyner, Harry (1977, p. 71) Magritte: Ideas and Images.

[18] Even though mental imagery covers all senses, I will focus on the visual modality since most evidence corresponds to it and visual mental imagery research dominates the literature.

in which both internally and externally generated images are projected to a single visual buffer for posterior inspection.

1.4.2. Kosslyn's visual buffer

According to Kosslyn's model of imagery, both bottom-up visual perception and top-down visual mental imagery are projected down the same visual pathways onto the same visual buffer used for object recognition: "It is as if the visual buffer is a kind of screen, which can display input from a camera (perception) or videotape (imagery)" (Kosslyn & Shin, 1991, p. 529). The visual buffer is the medium through which images occur: both internally generated images and visually encoded percepts are projected onto it.[19]

Neuroimaging research supports this claim showing that mental imagery draws on much of the same neural machinery as perception within the same modality (Dijkstra, 2019; Kosslyn et al., 2001). Studies reveal that mental imagery engages primary visual areas—more specifically, areas 17 and 18, the first parts of the cerebral cortex to receive input from the eyes (Kosslyn & Thompson, 2003; Pearson et al., 2015; Sparing et al., 2002). This has led some to claim that mental imagery can function much like a weak version of afferent perception (Pearson, 2019).[20]

Kosslyn suggests that, once projected onto the visual buffer, mental images are inspected in the same way as percepts (Kosslyn & Shin, 1991, p. 530). More specifically, he claims that to identify the object projected onto the buffer, it is compared to associative memories that contain information

[19] Kosslyn proposes the primary visual area as the most likely neural substrate for the visual buffer (Kosslyn, 1994). This area is the first stage of cortical processing of visual information.

[20] I take Kosslyn's Visual Buffer to refer to the same as the "active blackboard", a metaphor used by neuroscientists to describe early vision (Bullier, 2001; 2004). According to it, early visual cortex function is like a blackboard. Importantly, both bottom-up retinal sensory stimulation and top-down generated mental imagery can "draw" on this blackboard (Nanay, 2021).

about shapes and parts of objects. If the new information matches a previously stored pattern, one can identify the object.

The link proposed by Kosslyn between the visual buffer and associative memories is in line with the idea of the *Prima Facie* system reacting similarly to precepts and mental images. As an extension of this logic, I propose that the visual buffer could project its images to the *Prima Facie* system regardless of their origin. Imaginings, stripped of information about the source generating them, would be *read* at face value—as if they came from perceptual experience. This would also be the case for perceived representations of experience (film, Virtual Reality, etc.), which would trigger associations as if they had originated directly in perceptual experience. Because the Prima Facie system is isolated from information about the source of the representation, emotional responses and associations can be triggered (e.g., emotional valences towards the real stimuli) and created or reinforced during experiential imagination (e.g., conditioned responses).

The *Prima Facie* view does not deny nor rule out the existence of higher-order metacognitive systems responsible for integrating source information (Dijkstra, Kok, & Fleming, 2022). Indeed, it is accepted that monitoring mechanisms are responsible for downregulating overall reactions at the personal level when imagining. Consequently, *prima facie* effects can be reduced by the posterior integration of information about the source of the representation resulting from the monitoring mechanism. For instance, this view is compatible with the possibility that the affective system receives information on the source of the representation at a later stage, hereby moderating its *prima facie* responses. For instance, imagining that a loved one dies usually causes emotional reactions that are later reduced when the affective system receives information on the source of the representation, hereby moderating its *prima facie* responses. Daydreaming is often easily interrupted by a simple, first personal observation that informs us that “this is not happening”, which automatically reduces the negative (or positive) emotional response elicited by the simulation.

The *Prima Facie* view supposes an advantage over the Overshooting theories, since it accommodates for empirical evidence without appealing to a direct link between imagination and belief, and can still explain why there are clear differences between imagination and perception. Associations are a plausible candidate for explaining the effects observed of imagination on attitudes and behavior.

1.4.3. Unity of the phenomena

One possible objection to the *Prima Facie* view is that the heterogeneity of the presented findings precludes them from being a unified phenomenon. However, one can find a crucial shared take-away within the broad spectrum of literature reviewed: Experiential imagination has analogous consequences to real experiences. Furthermore, these effects are not due to monitoring errors, since people correctly identifies the source of the imagining at the personal level. And yet, the imagining ends up playing a functional role similar to that of observed evidence, being subpersonally integrated as if it had evidentiary value.

It is important to note that besides the central tenets of the *Prima Facie* view on the implicit assertoric force of imagination, the phenomena reviewed seem to suggest 1) what I will call *disruptive associationism* and 2) an "empirical bias" in judging the source of our associations.

a) Disruptive associationism

In the Treatise of Human nature, Hume gave a detailed account of associationism as a theory of learning. His theory concerned how perceptions ("Impressions" fruit of our incursions into the world) determined trains of thought (successions of "Ideas"). He contended that if impressions (IM1 and IM2) were associated in perception, then their corresponding ideas (ID1 and ID2) would also become associated in the mind. Mandelbaum (2020) generalizes Hume's theory of learning in the following terms: "If two contents of *experiences*, X and Y, instantiate some associative relation, R, then those contents will become associated, so that future activations of X will tend to bring about activations of Y" (emphasis added). However, a unified explanation of the phenomena here presented amounts to acknowledging that ideas are not only associated based on the sequences of impressions, but also based on sequences of ideas (experiential imaginings). Paraphrasing Mandelbaum, this addendum could read as follows: "If two contents of *experiential imagination*, X and Y, instantiate some associative relation, R, then those contents will become associated, so that future activation of X will tend to bring about activations of Y". This last claim is nothing new under the sun: Many empiricists acknowledged that mere thought could cause associations: "when two impressions have been frequently experienced (*or even thought of*) either simultaneously or in immediate succession, then whenever either of these impressions or the idea of it recurs, it tends to excite the idea of the other" (Mill, [1843] 1963 p. 852, emphasis mine).

What I claim is not only that imagining creates associations between the contents entertained, but rather that the associations created are unavoidably generalized to the flesh-and-blood counterparts. By disruptive associationism I mean the following. Perceptual experiences create associations between ideas in our minds, but internally triggered imaginings can disrupt, interfere with, and modify the associations originated in perception. Because of the *prima facie* processing of mental images, we do not have a set of associations for *imagined John* and a different set of associations for *John himself* (namely, one set of associations for the internally generated mental image of John and one for perceived John). Associations formed or broken when imagining are automatically generalized to the flesh and blood counterparts of the mental images. For example, if idly I imagine John betrays me, stipulating in the imaginative exercise that the *imagined John* is just a potential version of him will be sterile in avoiding the strengthening of the association between the idea of treason and the idea of *John*.

b) The "empirical bias"

To account for the attitudinal and behavioral effects, a second addendum is in place. I name the "empirical bias" our tendency to assume—by default and from a first-personal perspective—that associations in our mind have been caused by experiences. Nevertheless, in fact, they can also be entirely caused by experiential imaginings. This, of course, is not necessarily accessible through introspection, but rather an important tenet of stored associations.

Associations created by imagination are treated as if they had been caused by real experience. This is the route by which associations created when imaginings can influence beliefs. Benoit, Paulus, and Schacter (2019), recently showed that imagining an event can change attitudes towards its constituent elements. In two experiments, participants were asked to imagine people they liked (or disliked) in a neutral place (e.g., a living room). Only by virtue of imagining, the neutral place acquired the same valence as the person that they imagined in that place. Merely imagining confers emotional valence to the place. This effect of imagination mirrors the process that takes place when experiencing a negative event in a neutral place. For instance, going through a traumatic event at a neutral place (e.g., a breakup at a city park) makes this initially neutral place inherit the negative valence of the event. Imagination can create mental experiences that leave a strong-enough trace in our affective system, and then impact how we react to new experiences. Importantly, we are incapable of disregarding this affective component based on the source (imagination): The internal, fictional experience has changed the interpretation of the external world. This is what I call the "empirical bias", namely, to act as if our associations were always grounded in reality (i.e., caused by perceptual experiences). Take the following example: I had a neighbor (Paul) who used to frequently use a shovel in the garden. He seemed to be an amazing night gardener, but my deliciously cynical flatmate invited me to entertain the idea, in imagination, that each time the neighbor was shoveling dirt, he was indeed burying a corpse. After entertaining this imagining several times with all the necessary details, the neighbor soon became negatively valenced to me—as empirical evidence predicts. A certain uneasiness began to take hold of me every time I passed him on the stairs. This probably influenced my estimation of him being a trustful person, or, if asked, my subjective probability estimation of him committing a murder. On the other hand, some of my behaviors were also influenced. Would I dare to ask him for an ingredient I am missing for a recipe on a Friday night? I would certainly hesitate, and nothing would remove the automatic negative feeling originated by the mere thought of a future interaction with him. The "empirical bias" is the intuitive reaction of grounding the origin of such feeling on an experience. This "empirical bias" can be sometimes introspected along the following lines: "If I have this feeling, it is because of something I have *experienced*; a previous perceptual experience has associated this feeling toward the neighbor when thinking about him." As we have seen, nonetheless, merely imagining an event can change our attitudes towards its constituents.

1.5. Alternative explanations

1.5.1. Spinozian models of belief formation

Before proposing the existence of a completely new system, it might be more parsimonious to borrow from our understanding of other processes that are also involved in the reviewed evidence, such as belief formation. The Spinozian model of belief formation (SBF hereinafter; Gilbert, 1991; Mandelbaum, 2014) proposes that propositions are processed as true previous to comprehension. The model holds that the activation of a mentally represented truth-apt proposition leads to immediately believing it

(Mandelbaum, 2014, p. 55). Even when a proposition is rejected, the mere entertainment of the proposition in a previous stage affects our future evaluation of the truth value of the proposition.

It is important to notice that mental imagery has no role in SBF; the theory is concerned only with propositional content[21]. In an attempt to adapt this model to the current set of findings, a Spinozian could claim that propositions are part of the content of experiential imagination since they determine the nonvisual aspects of what is represented and appeal to this propositional content to explain the effects of imagination. In other words, when imagining, subjects entertain—and therefore, temporarily believe—the proposition that the imagining represents. For instance, in imagining her interaction with a dog, subject with a phobia of dogs might be entertaining the proposition "*This* dog is harmless" or "Dogs are harmless". This would constitute a non-imagery account of the evidence reviewed; that is, one in which the imagery involved would have no role in explaining the consequences of engaging in imagination.

There is a crucial caveat in reducing the effects of experiential imagination to a Spinozian model of belief: The quasi-experiential character of mental imagery plays a crucial role in its behavioral and attitudinal effects. Patients with phobias do not improve by entertaining mere propositional information. In fact, this is the first approach in treatment: making sure that patients are aware of the irrationality of the phobia and that they agree that the fear-inducing stimulus is not highly dangerous. That is to say, the patient has entertained the proposition that dogs are harmless on many occasions. She has read it and heard it countless times from his relatives in the face of her avoidance behaviors. But this has not been enough to change her behavior, nor the emotions elicited by the stimulus. The experiential character of the exposure in the imagination (via mental imagery) is essential.

[21] In Mandelbaum's words (2014, p. 61): "People do not have the ability to contemplate propositions that arise in the mind, whether through perception or imagination, before believing them. Because of our mental architecture, it is (nomologically) impossible for one to not immediately believe propositions that one tokens".

The following example by Dadds et al. (1997), about a case of claustrophobia associated with elevators exemplifies this:

> "His irrational fear remained that the world could come to an end when he was trapped in an elevator, and he would thus die, trapped there alone. Attempts to deal with this fear with rational countering (i.e., propositional cognition) were doomed to failure because it had to be conceded that, despite it being incredibly unlikely to happen, yes, the world could end with him trapped in an elevator. Adoption of a representational approach alerts the clinician to deal with the image itself, to attack its power to distress the person." (Dadds et al., 1997, p. 101).

Additionally, a propositional account cannot explain the fact that effects are modulated by the level of detail or vividness of visual imagery. For instance, the Imagined Contact effect has been shown to be enhanced by detail and the reduction of sensory perception. Husnu and Crisp (2011) showed that participants who were asked to generate more detail in their simulated encounter had higher expectations of having a greater number of out-group acquaintances in the future. Also, participants instructed to close their eyes during an imagined encounter had subsequently higher intentions to engage in future contact with outgroup members.

Finally, outside the visual modality, SBF also fails to explain the effects of second versus first person self-talk. If the effects were merely because the subject entertains a proposition and in doing so inevitably believes its content, the effect should be the same in both cases (e.g., "I can do it" versus "You can do it"), since once indexicals are disambiguated, the proposition is the same ("subject X can do Y"). Nonetheless, there are differential effects in holding one or the other. The *Prima Facie* view can account for them by appealing to the *Prima Facie* character of auditory imagery in the associative system and to the second person being associated with external judgements.

1.5.2. Gendler's notion of *alief*

The evidence here presented goes in the same direction as the examples introduced by Tamar Gendler (2010) in her description of Imaginative Contagion. She described this phenomenon as the cases in which imagining or pretending P has effects that one might expect would come only from believing or perceiving P.

One of the examples she gives is the bystander apathy effect—a well-documented phenomenon in social psychology. This effect concerns the fact that if one believes one is alone when presented with another subject in distress, one is likely to provide help more quickly than if one believes others are also present. Gendler echoes evidence showing that this effect takes also place if we merely imagine being in a group (Garcia et al., 2002, p. 845). In giving an account of these effects, Gendler mentions the source-indifference of some operations in processing imaginings. Nevertheless, she ends up interpreting imaginative contagion as a particular case of the mental state of *alief* (Gendler, 2010, p. 275). While I cannot raise here a detailed critique of the notion of *alief* she postulates[22], my aim will be mainly to show that, while it is well-equipped to account for synchronic effects of imagination, it cannot account for more diachronic effects. Gendler defines *alief* in the following way (2010, p. 255):

> "…an *alief* is a mental state with associatively linked content that is representational, affective, and behavioral, and that is activated—consciously or unconsciously—by features of the subject's internal or ambient environment. It is a more primitive state than either belief or imagination: it directly activates behavioral response patterns (as opposed to motivating in conjunction with desire or pretended desire)."

[22] For a critique on this notion, see Mandelbaum, 2013.

Gendler uses *alief* to cover a variety of cases in which our behavior departs from our professed beliefs. For example, stepping onto a safe transparent surface at a height can induce feelings of vertigo. In this scenario, she claims that although the adventurer *believes* that the walkway is completely safe, she also *alieves* something different. The *alief*, in this case, has the following content (2010, p. 256): "Really high up, long, long way down. Not a safe place to be! Get off!!" In Gendler's View, the effects of Imaginative Contagion would be due to *aliefs*. My main critique is that Gendler's account cannot explain the diachronic effects of imagination. *Alief* seems to be a powerful tool in accounting for behavioral responses while (or right after) imagining, but it cannot account for changes in the epistemic status of the subject caused by imaginings. For instance, it cannot explain the influence of imaginings on future probability estimations, nor the enduring changes in behavioral responses after imaginal exposure. Besides the considerations on the diachronic effects, positing a new mental state subtracts parsimony from Gendler's view when compared to the *Prima Facie* view, especially considering that Gendler also makes use of the notion of association in defining *aliefs*.

1.6. Objections to the *Prima Facie* View

1.6.1. Functional concerns

At this point one might be asking about the adaptive usefulness and function of the *Prima Facie* system. It is worth emphasizing that although the effects of the described *Prima Facie* system can sometimes be maladaptive, advantages of a direct link between imagination and the associative system are also worth considering.

First, the associations triggered by imagining could be useful in affective forecasting. In simulating a prospective event, the *prima facie* processing of the imagining triggers affect similarly to the corresponding experience, which may improve our capacity to anticipate how we would feel in that situation. It seems reasonable to postulate that for an anticipated event to affect us, it must be processed, in some sense, as *factive* or occurrent. The *Prima* Facie system would serve this function. The triggering of associated affect when imagining can also be adaptive in the following sense: In hard situations, such as Fred's example at the beginning of the chapter, imagination can help deal with frustration by representing a desired outcome (e.g., finishing a chapter) and the emotions associated with it (e.g., happiness). Additionally, the fact that imagining leads to the creation and reinforcement of associations may have several benefits. First, it makes it possible to be conditioned by experientially imagining content accessed through testimony (and not through direct experience). For example, if a friend reports that she had been bitten by the neighbor's dog, I might imagine the scene, which causes the image of the dog to acquire a negative valence. This, in turn, may be beneficial by adapting my subsequent behavior when seeing the dog accordingly. Second, the *Prima Facie* system can generate associations when engaging in regretful counterfactual imaginings about what could have been done, making these associations available for future use. For example, Sally could regret not asking for a salary increase in her last meeting with her boss. By repeatedly imagining what she should have said, she can create accessible associations between her boss's possible responses and her reactions to them, facilitating future action.

1.6.2. Expected effects

A resistance to embrace the *Prima Facie* view could also be due to the apparent lack of everyday evidence of the phenomenon described. Such an objection could be formulated as follows: if imagination had this functional profile, it would have bigger effects than it does on our daily life. From this point of view, the evidence presented here would thus be a rare and decontextualized laboratory/clinical phenomenon. Making use of intuition—as the one who raises this objection would do—I proceed to consider everyday phenomena that resemble the evidence reviewed and that could be explained by the *Prima Facie* view.

It is reasonable to think that the effects reflected in the empirical evidence would occur in everyday life when we imagine people, or ourselves, doing certain things. For example, the other day a friend contemplated the possibility that her partner was being unfaithful. The more she immersed in and elaborated on this possibility, the more she became convinced that her partner had been unfaithful. The imaginings, even correctly monitored, seemed to play an evidentiary role in the genesis of her conviction. Skeptics can try this simple exercise: Vividly imagine, for a while and repeatedly during a week, your neighbor burying a corpse (as I did thanks to my nice roommate). Then, ask yourself if you would leave your children in the care of this same neighbor. If the answer is *yes*: Do you anticipate that doing so would be safe? *Prima Facie* theory postulates that the neighbor will be associated with the imagined crime and acquired a negative valence. Because the associative system does not track the source of imaginings, these associations are going to be triggered when thinking about our neighbor after the imaginative exercise, which might influence our judgements about and our behavior towards him.

A closer look provides plenty of *Prima Facie* effects in everyday life. Take, for example, Martha Stewart. She recently revealed on an interview that she had to break up with Anthony Hopkins after he starred in the thriller *The Silence of the Lambs*[23]. Stewart said: "I have a big, scary house in Maine that's way by itself on hundred acres in the forest, and I couldn't even imagine taking Anthony Hopkins there. I couldn't—all I could think of was him eating, you know…"—referencing Lecter's culinary habits. It turns out that she could not separate Hopkins from his character—the cannibal Hannibal Lecter. So, despite having full awareness that his partner was playing a role, the image of Hopkins was immediately associated with the atrocities committed in the film and, therefore, with the fear response that these atrocities trigger.

[23] "Ellen DeGeneres" show (20/1/2022).

Although at a personal level Stewart was aware that she was witnessing a movie and its contents were not world-sensitive, her associative system did not distinguish between images that track the world and images that represent it (in the movie). Therefore, these associations were created and later unavoidably extrapolated, via generalization, to the flesh-and-blood Hopkins.

In relation to the evidence regarding self-fulfillment fantasies, similar phenomena often occur around us, although they may go unnoticed. For example, on numerous occasions, we meet someone with a very high confidence in being able to achieve a certain goal, despite having done nothing of the sort in the past. The following question arises: Where do they get this certainty from? A feasible answer is that they have imagined themselves succeeding at the task, and that the ease of imagining it has been extrapolated to the ease of achieving it. Furthermore, the view asserts that fantasies about oneself can affect one's self-concept by creating associations which, in turn, influence one's behavior. The writer Steinbeck seems to point to this phenomenon when claiming: "Socialism never took root in America because the poor see themselves not as exploited proletariat, but as temporarily embarrassed millionaires." [24] If we repeatedly imagine ourselves in an idealized way or an idealized future (as Fred does in Walton's example at the beginning of the chapter), our self-concept may become based on our imagined self. This could explain why we sometimes act and defend the interests of our imagined (potential) self rather than the interests of our actual self—we identify ourselves with our imagined self. Returning to the first example, Fred's recognition that his richness is just fantasy does not guarantee that his fantasizing is epistemically innocuous for him. Repeatedly imagining this might influence his perceived likelihood of his life radically changing from one day to the next, the effort required for this to happen, or his self-concept and self-esteem. This, in turn, might influence his behavior by, for instance, preventing him from unionizing. Being aware of the

[24] As quoted by Ronald Wright in "A Short History of Progress" (2005 p.124)

insidious ways in which imagination can influence our attitudes about the world despite our recognition of imaginings as such would be crucial in preventing some of the epistemically undesired consequences reviewed here. Just as being exposed to information can influence our beliefs even when we know this information is false (Fazio, Perfors, & Ecker, 2020), seemingly harmless internally generated imagination can also have undesired epistemic consequences. Against romanticized views of the virtues of imagination, the reviewed evidence suggests that, from a strictly epistemic point of view, it is wise to be cautious if not moderate in indulging in imaginative exercises, even when we correctly monitor them at the personal level.

1.7. Conclusions

Evidence shows that experiential imagination parallels some of the attitudinal and behavioral effects of real experience. This happens in the absence of reality monitoring errors and in non-epistemic uses of imagination (that is, when subjects do not attempt to obtain any knowledge from imagination). The *Prima Facie* view accounts for them by claiming that experiential imagination has implicit assertoric force. According to this view, in certain subpersonal processes—associative and affective—the parade of contents involved in imaginings are treated undistinguishably from world-sensitive precepts. I have motivated the claim that engaging in imagination is, *de facto* and by default, far from epistemically innocuous. Besides straightforward belief formation, quasi-sensory imagination seems to have more insidious ways of influencing attitudes and behavior that demand further investigation. This chapter is far from providing a complete theory of imagination. What I have proposed is a plausible and parsimonious way of rethinking the functional profile of experiential imagination to account for evidence of the attitudinal, emotional, and behavioral consequences of experiential imaginings.

Chapter 2

You *can* get some satisfaction! Imagination, symbolic action, and symbolic satisfaction

Abstract. In this chapter, I give a novel account of intrinsic symbolic actions which involve inanimate objects and are done apparently for no further end. These appear in the analytic literature as counterexamples to the Humean model of action rationalization, in which they are explained by a desire and a means-to-end belief. To provide a satisfactory explanation of intrinsic symbolic actions, authors have appealed to emotions (Husrthouse, 1991; Smith, 1998), imaginings (Goldie, 2000), and the phenomenon of redirected responses in the animal realm (Kovach & De Lancey, 2005; Scarantino & Nielsen, 2016). My account combines Goldie's appeal to the imagination with Scarantino and Nielsen's appeal to displaced action tendencies. Symbolic actions are symbolically displaced imaginings. At their core, these actions carry frustration concerning the impossibility of acting in the grips of an emotion. In performing them, two phenomena occur synchronically. First, thwarted action tendencies are displaced in a non-arbitrary way and released. Second, while displacing such action tendencies the subject imagines she is performing the denied action. The release of these tendencies on an object symbolically related to the object that causes the emotion provides *sui generis*, symbolic satisfaction.

2.1. Introducing the phenomenon

> "Burning in effigy. Kissing the picture of a loved one. This is obviously not based on a belief that it will have a definite effect on the object which the picture represents. It aims at some satisfaction, and it achieves it. Or rather it does not aim at anything; we act in this way and then feel satisfied"
>
> L. Wittgenstein, 1979, p. 4.

Actions are often referred to as *symbolic*. From kissing a photograph to burning a national flag, many actions fall under this umbrella of symbolism. Even actions that we usually do not conceptualize as *symbolic* have indeed a similar nature, like bringing flowers to a mausoleum. All these actions share a common attribute: The subject performing them, and sometimes also the audience, takes the symbolic object to stand for an absent object. For example, the photograph of a beloved stands for the person and the mausoleum stands for the deceased. Symbolic actions are even more puzzling under Hume's classical theory of actions, which explains actions by attributing the agent a desire and the belief that the action is a means to an end: I go to the forest because I want to breach fresh air and I believe that there I will breathe fresh air. How can symbolic actions be explained under this theoretical framework? What are the means-to-end beliefs guiding those actions and what are the agents' motivations for doing them? Why do humans carry out actions involving inert things, such as caressing a dress of someone we loved or gouging holes in the picture of someone we hate? To what kind of states—beliefs, desires, imaginings, emotions—do we need to appeal to render such actions intelligible? And what kind of satisfaction do we obtain from doing them, if any?

Debates on the topic of symbolism have been mainly patrimony of continental dissertations (Cassirer, 1944; Langer, 1942) and psychoanalytic theory (Freud, 1916; 1917; Petocz, 1999). For example, psychoanalysis emphasizes the unconscious nature of the symbolic process by taking the symbol to be an unconsciously produced substitute. Dreams, in psychoanalysis, symbolically satisfy unfulfilled desires (Freud, 1914). It would be difficult to translate the dogmas of such traditions into analytic terms. My aim here is more modest. I will give an analytic and clear account of symbolic actions. I will start by reviewing the little analytic literature that directly addresses them.[25] Symbolic actions entered the debate tangentially as instances of emotional actions (actions done in the grip of an emotion; Goldie, 2000; Hursthouse, 1991; Kovach & de Lancey, 2005; Scarantino & Nielsen, 2016). After characterizing the phenomenon and sketching the desideratum for a theory of symbolic action, I proceed to show that neither a Humean explanation in terms of a belief-desire pair nor the mere appeal to emotions render these actions intelligible. I then proceed to evaluate Goldie's (2000) account appealing to the imagination (section 3) and Scarantino and Nielsen's (2016) emotionist model, in which symbolic actions are analogous to redirected actions in the animal realm (section 4). In light of the criticisms raised, I then formulate my own account (section 5), in which these actions are symbolically displaced active imaginings.

2.2. What should an account of symbolic actions explain?

How simple would life be and how few therapists would be needed, if subjects restricted themselves to performing the actions philosophy has used as paradigmatic exemplars. Switching the light on and off: a pristine belief-desire pair always at hand to explain and motivate the action. However, bizarreness abounds, and such pristine explanations often shed scarce light on *why* an action was performed. This is the case of many actions in which inanimate objects are involved. It is frequent in colloquial language to refer

[25] An analytical examination of this type of actions will however raise epistemic concerns on practices such as psychoanalysis, which attribute the realization of unconscious symbolic actions to the subject.

to such actions as *symbolic*. Rather than begin with definitions, I will begin with exemplars—ideal cases that serve as models of the actions in question. Consider the following:

[1] Rolling around in one's dead wife's clothes out of grief (Hursthouse, 1991)

[2] Gouging holes in someone's picture out of hatred (Hursthouse, 1991)

[3] The ceremony of *burning someone in effigy* because they fled from justice during the inquisition (Wittgenstein, 1979)

[4] Prison guards that, after beating captives to death, continue to beat them (Postfunctional action: Kovach & De Lancey, 2005)

[5] Kissing a flag.

[6] Pummeling a cushion out of hatred while imagining it to be one's bank manager (Goldie, 2000).

[7] X arguing with Y, in virtue of X unconsciously identifying Y with Z (Unconscious identification: De Sousa, 2007).

A common feature of all these actions is that an object stands for another in them. In [1], the dress stands for the wife; in [3], the effigy stands for the person condemned. In symbolic actions, objects stand in a relation of representation with an absent object (e.g., a subject in [1], an abstract ideal in [4]). The choice of the representational object is not arbitrary, and a theory of symbolic actions needs to account for the conditions that allow for one object to stand for another. In those actions, the object of the symbolic action instantiates an associative relation with the object represented—e.g., contiguity in [1] and [4] and similarity in [2] and [3].

Symbolic actions, therefore, have a dual nature: An absent object is represented, and a present object stands for it. As Goldie puts it: "The symbolic nature of the expression takes place as it does partly because the literal action, as it were, is not a realistic option" (2000 p. 29). Because the relationship between the symbolic and the represented object is not arbitrary, the meaning of symbolic actions goes beyond external appearances. Symbolic actions can be thought of as analogous to metaphors: In the same way that a metaphor cannot be reduced to the literal meaning of the words, the significance of symbolic actions cannot be merely reduced to externally observable facts. As Skorupski (1976) indicates, symbolic actions have an analogous structure to the action that is represented when performing it:

> "The symbol substitutes for the thing symbolized (…) it is treated for the purposes of symbolic action as being what is symbolized. On this picture, the structure of symbolic action is clear: it represents or enacts an action, event, or state of affair in which the thing represented by the symbol plays a part analogous to that which the symbol plays in the symbolic action itself." (Skorupski 1976, p. 123)

The subject of [1] is not merely gouging holes in a picture, and the subject in [3] is not just caressing a dress: To properly understand what the subject is doing we need to know the symbolic meaning of the object of the action in each case, in virtue of whether it is standing in for another object an, if so, which. Clarifying the notion of symbolic meaning is therefore a desideratum of a theory of symbolic actions.

Last, but not least, in symbolic actions subjects perform an action analogous to one they cannot perform. The action seems to be a substitute: The grieving widower would not roll around his wife's clothes if he could instead hug her. Because of the non-arbitrary relationship between the symbolic object and the absent one, the action seems to provide satisfaction *similar* in some sense to the unavailable courses of action. The satisfaction provided by symbolic actions is not merely alternative to the action that cannot be performed (as hugging a friend would be in the case of the widower). The similarity between the satisfaction provided by the symbolic action and the unachievable satisfaction also needs to be addressed by a theory of symbolic action.

To summarize, a theory of symbolic action should: 1) explain why subjects perform intentional actions and render them intelligible in light of such explanation 2) specify the relationship between the symbolic object—its *symbolic meaning*—and the represented object, and 3) clarify the kind of satisfaction subjects obtain from performing those actions. In the following section, I argue that, concerning desideratum 1, neither appealing to belief-desire pairs nor simply mentioning emotions is enough to render intrinsic symbolic actions intelligible as well as the motivation of agents in doing them.

2.3. Why do we perform symbolic actions?

The orthodox view in the philosophy of action claims that a belief-desire pair always explains intentional action. According to the standard Humean model, a belief-desire pair causes and explains any intentional action (Smith, 1998; Davidson, 1963). The Humean model can be generalized in the following way: For any intentional action A, the agent did A because she *desired* to X and *believed* that doing A would be a means to X.[26]

In this account, Fred's intentional action of turning the light on is explained by his *desire* to illuminate the room and his *belief* that he would do so by turning the light on. Some—the less interesting tokens—of action types [1]-[7] are amenable to this explanation. I will name these *instrumental symbolic actions*—symbolic actions which can easily be characterized as a means to an end—, in contrast to *intrinsic symbolic actions*—symbolic actions that cannot be characterized as a means to an end. In this chapter, I will only be concerned with the latter symbolic actions. Next, I motivate the distinction between *instrumental* and *intrinsic* and the plausibility of intrinsic symbolic actions.

[26] I borrow this formalization from Scarantino & Nielsen, 2016.

2.3.1. Instrumental versus intrinsic symbolic actions

Instrumental symbolic actions can be divided into two groups: communicative and superstitious symbolic actions. Here is an example of communicative symbolic action: Alba, a tennis player, kisses her husband's ring after winning a set to show him—in the audience—gratitude for his support. In this case, the ring "stands" for the husband, and the agent uses the audience's awareness of this relation to communicating something. A belief-desire pair explains the action and renders it intelligible: The agent kissed the ring because she *desired* to communicate gratitude to her husband and *believed* that kissing the ring would be a means of doing so.

On the other hand, the other set of instrumental symbolic actions responds to superstition or religious belief. This is the case for some tokens of action type [3]. During the Spanish Inquisition, it was frequent for people to be judged *in absentia.* If the convicted fled from justice and thus avoided a death sentence, it was common to burn an effigy of her in the village square. The effigy symbolized the fugitive. The "executors" and the audience believed that burning the figure was causally efficacious in damaging the fugitive, wherever he or she was. The action, therefore, was perfectly amenable to a Humean explanation: They burned the effigy symbolizing the fugitive because they *desired* to hurt him or her and *believed* that burning the effigy would do so. A whole range of other symbolic actions is also caused by superstitious or religious beliefs. Such is the case of the Ushbeti, small statues Egyptians placed in their relatives' coffins believing that they would transform into slaves in the afterlife. It is also the case of African ceremonies in which pins are stuck into Voodoo dolls, guided by the belief that this will cause ailments in the parts of the subject represented by the doll.

However, different actions tokens of the same type (e.g., kissing a ring or burning an effigy) can be symbolic in the intrinsic sense that concerns us. More specifically, they are pursued for their intrinsic value instead of as a means to an end. For instance, in the case of people judged *in absentia*, many times the accused died before the trial finished. Still, he was burned in effigy.

In these cases, positing an instrumental motivation for the action seems out of place. It is absurd to postulate any belief on the action being causally efficacious in hurting the person, who has already died. Beating up and jumping above statues of dictators after their fall has also brought pleasure to discontent citizens throughout history, even after the death of the dictator in question. Similar to burning an effigy after the fugitive has died, we can interpret some cases of the so-called *postfunctional actions*—actions that continue even after their ostensive goals have been achieved—as instances of intrinsic symbolic actions. Consider [4]: During the war in Bosnia, reports surfaced about prison guards beating captives to death and then continuing to beat their corpses until they fully disintegrated. Concerning this postfunctional action, Kovach and DeLancey write: "Nevertheless, at the time of the beatings, the guards may have been motivated primarily by intense rage, which did not abate until long after the victims had died" (2005, p.119). I take postfunctional actions as instances of intrinsic symbolic action. The enemy—no longer alive, no longer *hurtable*—is symbolized by its corpse. The action, in this case, no longer has the end of hurting the prisoner. One last instance of intrinsic symbolic action worth mentioning concerns [1]: The grieving man who "takes his dead wife's clothes out of the wardrobe, puts them on the bed and rolls in them, burying his face in them and rubbing them against his cheeks", but has no further purpose in doing so, as described by Hursthouse (1991, p. 59).

In the case of intrinsic symbolic action, it is difficult to spell out the agent's reasons in a way that makes the behavior intelligible to us. What is the agent's aim in these actions and why is this aim important to him? In the following section, I will focus on intrinsic symbolic actions, that is, symbolic actions made for no other end and in the absence of communicative purposes and superstitious or religious beliefs. I will refer to them as symbolic actions throughout the paper. Let us now see whether a belief and a desire held by the agent can render intrinsic symbolic actions intelligible.

2.3.2. Hursthouse's emotionist model

Rosalind Hursthouse (1991) argued that certain actions done in the grip of emotion are not satisfactorily explained by mentioning a belief-desire pair held by the agent. She named these actions *arational actions* since subjects seem to perform such actions for no reason. According to Hursthouse, arational actions elude an explanation in Humean terms. The set of actions she described was quite heterogeneous from rumpling someone's hair (out of love) to jumping up and down (out of joy) or gouging holes in someone's picture (out of hatred). Several instances of what Hursthouse defined as *arational actions* are also instances of intrinsic symbolic actions. This is the case for the following two (corresponding to [1] and [2] above):

> [1] "…tearing one's hair or clothes, caressing, clutching, even rolling in, anything suitable associated with the person that is the object of grief, e.g., pictures, clothes, presents from her…" (1991, p. 58).
>
> [2] "…Jane, who, in a wave of hatred for Joan, tears at Joan's photo with her nails, and gouges holes in the eyes" (1991, p. 59).

According to Hursthouse, these intentional actions are not means by which agents realize their goals, and therefore are not explained by a belief-desire pair. Looking for an explanation of people's reasons for performing these actions, Hursthouse evaluated two potential candidates from a Humean stance. The first belief-desire pair she considers is the desire to express an emotion and the belief that by doing A, the emotion will be expressed. The second Humean pair she evaluates is the desire for pleasure and the belief that doing A would bring pleasure. She discards these options because of the implausibility of ascribing someone who is grieving for his dead wife a belief that he simply need not have. Hursthouse concluded that no belief-desire pair can ground a Humean explanation of these actions. Instead, she claimed that such actions were better explained by merely mentioning the emotional state of the subject, which alone is a sufficient explanation for the action:

> "On the very many occasions on which such actions were performed it would be true to say ... : (i) that the action was intentional; (ii) that the agent did not do it for a reason in the sense that there is a true description of action of the form 'X did it (in order) to ...' or 'X was trying to ...' which will reveal the favorable light in which the agent saw what she did and hence involve, or imply, the ascription of a suitable belief; and (iii) that the agent would not have done the action if she had not been in the grip of whatever emotion it was, and the mere fact that she was in its grip explains the action as much as anything else does" (Hursthouse 1991, p. 59).

Her explanation of why Jane gouged holes in Joan's picture [2] is that she was in the grip of hate, and, because of this, she desired to gouge holes in his picture. Hursthouse's model for intrinsic symbolic actions can be formalized in the following way:

- **Hurtshouse's emotionist view:** For an intrinsic symbolic action A, the agent did A because she was in the grip of some emotion E and desired to do A.

Before raising criticisms of Hursthouse's account, it is worth seeing Smith's response to it from a Humean stance.

2.3.3. Smith's Humean account

Defending the Humean model from Hursthouse's attack, Smith argues that Hursthouse has ignored the most plausible belief-desire pair Humeans could appeal to in explaining actions such as [1] or [2]. Smith explains [1] in the following way: The widower did A because he had the *desire* to roll around in his dead wife's clothes and the *belief* that he could do so by rolling around in those clothes that he rolled around in. Surprisingly, by this explanation, the ends and the means are one and the same. That is, in the belief that explains the action, the means are equated with the ends. Smith himself acknowledges that this Humean account is "distinctly unsatisfying" and that it provides us "with very little illumination" (1998, p. 160). The account is unsatisfactory because the belief-desire pair leaves the central question unanswered: Why would anyone want to roll around in his dead wife's clothes, as in [1]? Or why would anyone want to gouge holes on a piece of paper, as in [2]? In dealing with this question, Smith provides a supplement for the Humean account. He claims that the origin of such bizarre desires can be illuminated by appealing to the emotional state of the subject. In the context of grief, he claims, having this desire is completely normal:

> "…grief at the loss of a loved one is, by definition, a state in which we are disposed to think, and to desire, and to do, all sorts of things: cry, dwell on memories of the loved one, seek out things that remind us of the loved one and hold them close, and so on and so forth" (1998, p. 160).

Appealing to grief, Smith claims, makes the origin of the belief-desire pair that motivates [1] intelligible. By Smith's account, the motivation (a belief-desire pair) is the central element explaining intrinsic symbolic action; The emotion only complements the explanation by shedding light on the causal history of this motivating reason. Smith's model can be generalized in the following way:

- **Smith's Humean model:** For an intrinsic symbolic action A, the agent did A because he was in the grip of an emotion which caused him to *desire* to do A and *believed* that doing A would be a means to A.

Both Hursthouse's emotionist account and Smith's Humean account are unsatisfactory. Hursthouse's appeal to emotion to explain the action and Smith's appeal to emotion to explain the origin of the belief-desire pair leave the agent's desire unexplained. Why would grief give rise to the desire to caress a dress or hate to the desire to make holes in a specific picture? Mentioning the emotion is unhelpful in explaining why in grief or anger people want to perform these kinds of actions do this kind of thing. This concern of emotions not making these actions intelligible was raised by Goldie (2000). Goldie objects to Smith claiming that the desires posited to explain these actions are not "primitively intelligible" even in the presence of these emotions. Goldie defines as primitively intelligible those desires that cannot be explained in virtue of anything else other than the emotion of which they are part. Considering Jane's hatred, for example, the desire to scratch out Joan's eyes (the subject causing the anger) is primitively intelligible. On the other hand, the desire to scratch Joan's eyes *in a picture* is not primitively intelligible. Both Hursthouse and Smith restate the obvious fact that, while in the grip of the emotion, this kind of action appealed to the agent. However, what needs to be explained is *why* being in the grip of emotion makes the action appealing. This desideratum is formulated by Davidson (1963, p. 685; emphasis mine):

> "A reason rationalizes an action only if it leads us to see something the agent saw, or thought she saw, in his action–some feature, consequence, or aspect of the action the agent wanted, desired, prized, heled dear, thought dutiful, beneficial, obligatory, or agreeable. We cannot explain why someone did what he did simply *by saying the particular action appealed to him; we must indicate what it was about the action that appealed*."

Moreover, such accounts do not mention nor shed light on the relation between the object of the symbolic action and the one represented in the action. Although Hursthouse acknowledged the symbolic nature of many of the examples of arational actions, she gave no role to such nature in rendering the actions intelligible. However, the relationship between the object of the symbolic action and the object represented is rarely arbitrary, and this fact needs to be accounted for. Our third desideratum—explaining the kind of satisfaction obtained by the subject when performing these actions—is not mentioned in Hursthouse's or Smith's account.

For the reasons mentioned above, Hursthouse's and Smith's accounts are unsatisfactory when accounting for what motivates us to perform intrinsic symbolic actions. I turn now to Goldie's account, which introduces imagination into the explanation of symbolic actions.

2.4. Imagination enters the picture: Symbolic actions as an active pretense

Goldie (2000) distinguishes between two types of actions done out of emotion: reasoned actions—like jumping over a gate in fear of a bull—and genuine expressions—like lovingly stroking a face. The main difference between them is that genuine expressions of emotion are not a means to an end. Goldie takes instances of intrinsic symbolic actions to be genuine expressions of emotion and, therefore, done for no other reason than the expression itself. Because of this, he sees no problem in Smith's explanation of them in terms of a desire and a means-end belief in which the means and the end of the action are one and the same. That is, in [2] Jane gouges out the eyes in Joan's photograph because she *desires* to gouge out the eyes in Joan's photograph, and she *believes* that she can do this by gouging out the eyes in Joan's photograph.

However, Goldie acknowledges that, in Smith's explanation, the "bizarre" desire motivating the action is inadequately explained. He also claims that adding the supplement of the emotion (anger, in this case) does not explain the origin of the desire. In other words, even considering the emotional state of the subject, such "bizarre" desire is not primitively intelligible. In the case of anger, the desire to harm the object of our hate is also primitively intelligible. In the case of fear, the desire to be away from the object of the fear is primitively intelligible. However, Goldie argues—in contrast to Smith—, that the desire to scratch the eyes in a photo of a person with whom one is angry is not rendered primitively intelligible by merely mentioning that the subject is angry. Why would one desire, in anger, to do things such as scratch the eyes in a picture of the person causing the anger?

2.4.1. Goldie's account

Goldie's proposal to amend the Humean explanation is to appeal to *wishes,* a special type of desire. He defines *wishes* in the following way: "When I wish for something, I desire that thing, and I also imagine or am disposed to imagine, the desire to be satisfied" (Goldie, 2000b, p. 28). Goldie claims that, in Jane's example, what is primitively intelligible is her desire to hurt Joan. However, Jane does not act on this desire in the expressive action because of constraints placed by civilized society—and probably legal consequences. As a result, Jane only wishes to hurt Joan: She desires to scratch out Joan's eyes and she imagines satisfying the desire. Her action, therefore, is expressive of a wish:

> "By acting out through expressive action, Jane is, in a symbolic way, acting out just what she knows she ought not to do. So, this, I argue, is the way to make intelligible the 'bizarre' desire to scratch out the rival's eyes in the photo, by appeal to the intelligible desire, in the other vector, to scratch out the rival's eyes" (2000b, p.131)

In Goldie's model, Jane's desire to scratch Joan's eyes in a picture is rendered intelligible by further appealing to her primitively intelligible desire to hurt Joan. Goldie's account of intrinsic symbolic actions can be formalized in the following way:

- **Goldie's Humean model:** For an intrinsic symbolic action A, the agent did A because she was in the grip of an emotion, and she *wished* to do A' (she desired to do A' and was predisposed to imagining herself doing A'). As a result of this wish, she *desired* to do A (pretending to do A') and *believed* that by doing A she would be doing A (pretending to do A').

It is difficult to regard such a complex model as a mere amendment of the Humean account, as Goldie does. However, Goldie claims that his approach puts forth two distinct explanations which are not in competition. He claims that his account is Humean in nature since the belief-desire explanation (*à la* Smith) has a central causal role despite being supplemented by the appeal to emotions, other desires, wishes, and imaginings.

2.4.2. Why Goldie's model fails in explaining intrinsic symbolic actions

The main criticism of Goldie's view has been raised by Kovach and De Lancey (2005). It concerns the fact that not all imaginings of satisfying a desire lead to active pretending. In desiring to scratch Joan's eyes and being unable to do so, Jane could also just imagine that she scratches Joan's eyes out without an accompanying action. However, in this case, she engages in active pretense. As Scarantino and Nielsen point out:

> "Jane could easily imagine scratching out Joan's eyes and do nothing about it. It seems that in order to lead to action, Joan's wish would *need to cause an actual desire to scratch out the eyes in the picture*, and then combine with a belief that one can scratch out the eyes in the picture just by doing what one is doing" (2016, p.)

So, concerning the first desideratum for a theory of intrinsic symbolic actions—why do we do such *strange* things—, Goldie needs to explain why the desire to do something and the simulation of doing so brings about an action. Given that one could imagine without pretending, Goldie needs to account for why in such cases subjects do not merely imagine but also actively pretend.

Regarding the second desideratum—specifying the relationship between the represented object and the one present in the action—, Goldie's account does not specify the transition between the object that causes the primitively intelligible desire and the object of the symbolic action. Goldie notes that there is often a symbolic match between such objects. However, he claims that this need not always be the case and dismisses the explanatory relevance of this relationship. He provides the following example to motivate the irrelevance of the symbolic relation: While opening the mail during breakfast, you read that your bank is increasing their charges. As a response, you might imagine that a cushion is a bank manager and pummel it. Or, in anger, you may kick the nearest object (e.g., the kitchen table). In the first case, Goldie would explain the action in terms of a *wish* to hurt the bank manager and the tendency to imagine doing so.

According to Goldie the object of the intrinsic symbolic action can be chosen randomly, without any specific relation between it—the cushion—and the object causing the action—the bank manager. The agent can choose an object for the expression of emotion and imagine a relation to the representation. As it happens in pretense game, where we can stipulate that blue blocks may be considered 'cars' and red blocks 'trucks', in symbolic actions the agent need only behave "as if" an object represents another. According to Goldie, symbolic actions are cases of pretense, and the subject of the action can decide at will which object stands for the one causing the emotion.

I argue, in contrast to Goldie, that subsuming symbolic actions in pretense does not reflect the nature of the phenomenon, for two main reasons. First, the relation between the object of the symbolic action and the object it stands for is not arbitrary nor can it be decided at will. This central feature distinguishes intrinsic symbolic actions from pretense. Take the case of the [1]: In grief for the absence of his wife, the widower would probably not indistinctively caress his sister's dress instead of his wife's dress. It is implausible that merely stipulating that his sister's dress stands for his wife would invite him to caress it and roll around in it. In this case, the relation between the object involved in the action and the object causing the emotion is not arbitrary nor can be decided at will. Something similar happens in case

[2]. The satisfaction obtained by gouging holes in the picture of the person causing our hate would not be the same if, while hating him, we had a picture of our favorite writer in front of us and decided to gouge holes in his or her eyes while imagining that this picture stands for the hated person. Take also [4]: When feeling pride for whatever ideals his country represents, it would not be the same for the patriot to kiss his country flag or a random piece of white cloth that he takes to represent his country, in this specific action.

Against this observation, it might be argued that choices of representational objects are also guided by some kind of resemblance. For instance, because of its shape, a banana might stand for a telephone and a broomstick between the legs might stand for the mane and the back of a horse. I agree that this resemblance may be at play. However, the similarity is strictly functional, based on the shapes and uses of objects—e.g., one might not take a computer to stand for a horse. Nevertheless, this is not the kind of resemblance between the object of the action and the object represented in the action in the case of symbolic actions. In these cases, the functional resemblance between two objects—e.g., a Norwegian flag and an Icelandic flag—would not be enough for one to stand in for the absent object (the idea of a country) instead of the other in the symbolic action—kissing or burning the flag of one of these specific countries. The relationship between the object and its representation in symbolic actions is not random nor can it be entirely decided at will. In taking symbolic action as episodes of pretense, Goldie underestimates the relationship between the object of the action and the represented object, which in the cases at hand seems to elude arbitrariness or stipulation. The second difference between symbolic actions and pretense is that we usually pretend to some end. We might pretend that a broomstick is a horse to have fun with or entertain kids (Symbolic game: Vygotsky, 1933). In contrast, intrinsic symbolic actions do not seem to be means to an end.

Goldie does not explicitly address our third desideratum—explaining which kind of satisfaction subjects obtain by performing these actions.

Nevertheless, since he takes symbolic actions to be mere exercises of pretense, he implicitly puts forth that the kind of satisfaction obtained from these specific actions and objects (e.g., kissing or burning a specific flag or wedding dress) would differ from random objects (e.g., a piece of cloth) imagining they stand for abstract ideas, objects, or persons. His account does not leave room for the notion of a substitutive or symbolic satisfaction different from the one of pretense.

The discontent with Goldie's model invited authors to return to purely non-Humean, emotionist accounts of the symbolic actions, *à la* Hursthouse. In the next section, I address Scarantino and Nielsen's account (2016), which considers symbolic actions a subclass of emotional actions—namely, displaced emotional actions.

2.5. Symbolic actions as displaced emotional actions

Scarantino and Nielsen's account is partly inspired by Kovach and De Lancey's (2005) analogy between symbolic behaviors in humans and redirected behaviors in animals:

> "An example is the case of the lioness. Irritated upon having her tail bitten by a cub, she turns to a tree trunk and gives it a good scratching. What has happened is that an aggressive response to the cub has been redirected at another target. What works to spare the cub also spares the top of many a child whose exasperated parent pounds a tabletop instead. An explanation in terms of redirected anger makes what the lioness does intelligible. This suggests that there may be no need to invent cognitive epicycles to obtain an explanation of Jane's behavior. Unlike the lioness, Jane surely can make wishes, but like the lioness, she needn't" (2005, p. 119).

Ethology describes redirected responses in the animal world as "displaced activities" (Tinbergen, 1951). This notion was initially introduced to explain animal behaviors that appeared out of their usual context (Tinbergen, 1939). In several situations, the animal redirects a response caused by one object towards another. In the literature, these out-of-context behaviors are attributed to a conflict or a thwarted situation, in which the animal cannot obtain the desired satisfaction or satisfy a behavioral tendency. Ethologists (Lorenz, 1957; Tinbergen, 1951) have suggested that the function of this displaced animal activity is cathartic, "acting as an outlet through which surplus motivational energy could blow off" (Zeigler, 1940, p. 367). For example, birds exhibit displacement preening when copulation is prevented, i.e., when the situation is thwarted. Lions, for example, redirect aggression from a cub towards an inanimate object like a trunk in response to the conflicting motivation between the desire to bite the cub and the instinct to protect it. In their interpretation of symbolic actions as redirected responses, Kovach and De Lancey do not specify how redirection works in humans or whether it differs from animal responses. Their suggestion generalizes for intrinsic symbolic actions in the following way:

- **Kovach and de Lancey's model for intrinsic symbolic actions:** For any intrinsic symbolic action A under emotion E, the agent did A redirecting a response caused by X towards Y.

In displaced behaviors, Kovach and De Lancey do not distinguish between random redirection—like the parent pounding a tabletop (without taking the tabletop to stand for anything while pounding it)—and symbolic redirection—like Jane's gouging of eyes in a picture that stands for a person. However, there seems to be a difference between redirected actions in the animal kingdom and symbolically redirected actions in humans. Redirected actions in animals are frequently a common behavioral pattern within the species and, in the case of redirection of anger, for example, the nearest inanimate object seems to suffice. However, human symbolic actions are idiosyncratic, and as I have argued before, the symbolic object is not arbitrarily chosen nor determined at will. Scarantino and Nielsen's (2016) account attempted to amend this aspect by distinguishing between radically and symbolically displaced emotional actions. According to Scarantino and Nielsen, the goal of emotion can be pursued and satisfied, attenuated in virtue of the redirection of the attention to some other goal, or symbolically satisfied.

2.5.1. Scarantino and Nielsen's account

Following Kovach and de Lancey, Scarantino and Nielsen claim human symbolic actions are analogous to displacement in the animal realm. The analogy has to do with the causal origin of the action. According to Scarantino and Nielsen, civilizing constraints in humans are analogous to conflict situations in animals—as in the case of the lioness and the cub—, and the impossibility to perform a certain action is in humans analogous to the thwarting situation in the bird example above. The authors offer a very detailed account of emotional actions and consider symbolic actions a subgroup of these actions. Other emotionist accounts, such as Hursthouse's, use emotions to explain behavior without specifying a theory of emotions that marks that causal path. On the contrary, Scarantino and Nielsen make use of the Motivational Theory of emotions (MTE hereafter; Scarantino, 2014) to account for emotional actions. According to MTA, emotions are action control systems designed to prioritize the pursuit of certain goals over others (Scarantino & Nielsen 2016, p. 2897). Anger, for instance, has the relational goal of removing an object appraised as blameworthy. Crucially, MTA postulates a two-level control structure to account for emotional actions. The first level sets the goal of the emotion, and the second level exercises rational control over the goal—namely, it determines whether and how the goal settled by the emotion is pursued. Two phases are differentiated within this second level: a deliberative phase that tests the compatibility of the goal of the emotion with the emoter's other goals and values and an executive phase. As a final product of this process, the emoter can pursue the relational goal of emotion (e.g., jumping up and down out of joy), or the goal can be impeded in the deliberative phase if it is incompatible with other goals and values of the subject (e.g., being in public). If the latter is the case, the emotional action can be displaced.

Scarantino and Nielsen argue that the analogous conflict and thwarting situations in humans can lead to two types of displaced emotional actions: radically displaced and symbolically displaced. The difference between them derives from the kind of relationship between the object that causes the emotion and the object of the action: contingent or symbolic. Radically displaced emotional actions are those whose object does not stand either in an instrumental or symbolic relation to the object that caused the emotion. These displaced actions let the subject "vent" the emotion by redirecting attention away from its relational goal. Examples of this are adjusting one's tie out of fear before giving a talk or kicking the nearest object out of anger. Emotional actions can also be displaced symbolically. Scarantino and Nielsen describe symbolically displaced emotional actions as those "whose object stands in a symbolic relation to the object of the emotion that caused it" (2016, p 2994). An example of this could be [1] or [2]. Intrinsic symbolic actions bring, by this account, a symbolic satisfaction of the relational goal of the emotion. Gouging the eyes out of someone's picture allows for the symbolic satisfaction of the goal of anger, with is gouging a rival's eyes. Scarantino and Nielsen's account of intrinsic symbolic actions can be generalized in the following way:

- **Scarantino and Nielsen's model**: For any intrinsic symbolic action, the agent did S because she was affected by emotion E, which has X as its goal. X could not be pursued because of civilizing constraints or an impossibility and doing S was a way to achieve the symbolic satisfaction of doing X.

2.5.2. Why displacement is not enough

In the following paragraphs, I raise three criticisms of Scarantino and Nielsen's account. First, Scarantino and Nielsen justify grouping radically displaced and symbolically displaced emotional actions together by appealing to the fact that both classes of actions are analogous to displacement actions in the animal kingdom. Both kinds of actions, they claim, are the result of thwarting an action or a conflict situation. However, an important question remains unanswered: Why is the emotional action sometimes radically and sometimes symbolically displaced? For their account to be satisfactory it needs to be supplemented with an explanation of why the transition between the object that caused the emotion, and the object of the displaced action is sometimes arbitrary and sometimes symbolic. Is it related to the degree of concreteness of the object of the emotion? Elaborating on Goldie's example, maybe when receiving a letter from the bank and feeling hate the subject punches the table since he does not know who precisely the object of his hate is; but maybe when he perfectly identifies that the object of his hate is the director of the office, he prefers to engage in a symbolic game in the which he tears the director's business cards to shreds. It could also be that the kind of displacement depends on the emotion. For instance, can an action performed out of grief be radically displaced? Is it possible for a man remembering his late wife to caress the nearest object out of grief? Another missing clarification relates to the degree of urgency in deploying the action, which concerns the author's conception of emotions as prioritized action tendencies seem to be opportune. Do emotions involved in radically displaced actions have a *higher* priority tendency than the ones that go through a process of symbolization? Scarantino and Nielsen should specify why the transition between the object that causes the emotion and the object of the emotional action takes such a different route in each case. It is odd that in some cases the most proximal object suffices but, in the others, a symbolic process takes place. The complexity of the symbolic displacement is not accounted for in Scarantino and Nielsen's account, which speaks to the circumstances or reasons that invite radical or symbolic displacement.

Concerning the kind of relationship between the symbolic and the represented object in the action, Scarantino and Nielsen claim that we can imbue objects with symbolic value in a completely arbitrary way: "For a symbolic process to take place one must simply take a certain object to stand for another, perhaps only for a fleeting emotional action" (2016, p. 2992). The critique of Goldie in the previous section, therefore, applies to this account. As I have presented above, there are constraints on which object can stand for another, and paradigmatic symbolic actions do not seem to allow for arbitrariness or freedom in imbuing an object with symbolic value.

Concerning the last desideratum—explaining what kind of satisfaction we obtain from these actions—, the authors claim that we obtain *symbolic satisfaction* as the result of symbolically displaced emotional actions. However, they do not specify what they mean by this notion. What does it mean for something (e.g., a goal or a desire) to be symbolically satisfying? What is the difference between satisfaction and symbolic satisfaction? Why is one kind of satisfaction a symbol for other, unobtained satisfaction? Is the "symbolic satisfaction" obtained in symbolic actions a satisfaction concerning a state of affairs or satisfaction that results from imagination? These questions remain unanswered in their account, which specifies that symbolically displaced emotional actions bring about a specific kind of satisfaction, but does not explain its distinctiveness.

Scarantino and Nielsen's account is unsatisfactory in these three respects. In the next section, I will introduce a tentative account of symbolic action, which hybridizes Goldie's account with Scarantino and Nielsen's. I argue that components of those accounts, supplemented by the notion of *symbolic meaning*, constitute a promising candidate for addressing the essential questions a theory of symbolic actions must answer.

2.6. Intrinsic symbolic actions are symbolically displaced imaginings

In this section, I offer a novel account of intrinsic symbolic actions. This account hybridizes Goldie's appeal to imagination and Scarantino and

Nielsen's appeal to displaced emotional actions. In my account, symbolic actions are symbolically displaced imaginings. This account inherits Goldie's appeal to the imagination to account for the relationship between the symbolic and the represented (in the action) object, and Scarantino and Nielsen's appeal to displaced action tendencies to account for why the imagining is acted on. In this account, the symbolic action has a teleology but is caused non-rationally (i.e., is not performed for a reason, standardly understood as a belief-desire pair). I will present the account by answering the three main questions an account of intrinsic symbolic actions should address, namely:

1. Why do we perform those actions?
2. What is the relationship between the object of the symbolic action and the object represented in the action?
3. What kind of satisfaction do we obtain in performing those actions?

2.6.1. Why do we perform those actions?

Regarding the origin of intrinsic symbolic actions, I agree with what has been previously put forth by other authors: Symbolic actions have their origin in the impossibility of carrying a desired action out of emotion. In [1] the desire to hug a beloved one, out of love; in [2] the desire to act violently against the person depicted in the picture, out of hatred. The impossibility at the core of other symbolic actions may also be the desire to express affection towards abstract entities. Thus, affection or hatred towards the ideology of a nation may lead to kissing or burning a flag. I do not find it relevant to taxonomize the catalog of frustrations that prevent us from consummating an action (civilizing constraints, impossibility, et cetera). For our purpose, it is enough to acknowledge that intrinsic symbolic actions are primarily caused by the frustration of not being able to act in the grip of an emotion.

When we are unable to carry out the desired action, one way to deal with frustration is to imagine doing the action (e.g., the widower will daydream about hugging his wife, the angry person about acting violently against his rival). In intrinsic symbolic actions, two factors are in place: the imagining of performing the action and emotional action tendencies. While imagining doing X with object A, the action tendency (doing X) is symbolically displaced towards object B. The action is displaced towards this B in virtue of B's relationship with the object causing the emotion (A). B is a good stand-in for A in the action if it triggers associations that complement and increase the vividness of the imagining that is taking place (doing X with A). This will be clearer with an example. The widower in [1] might imagine hugging his wife out of grief. If when he does so he is walking down the street, he will not stop to hug a traffic light, since the traffic light does not improve his imagining of hugging his wife. However, if he is imagining hugging his wife in the privacy of his home and sees her dress, this object will complement and increase the effect and the vividness of the imaginative project he is immersed in. The dress will trigger memories of his wife wearing it, which will increase the emotional content of the imagining. Touching the dress will bring back memories of touching it on his wife's body. Perhaps, the smell will evoke the memory of a certain perfume. Since the object has this capacity to increase the detail and vividness of the imagining in which a frustrated action is evoked, it will increase the emotion and with it its action tendencies (hugging the woman). The co-occurrence of an imagining of doing X with A, the emotional action tendency of doing X, and the presence of an object B, will lead, in the appropriate circumstances, to doing S with B (e.g., rolling in the dress in [1]).

This explanation complements the questions unanswered by Goldie's and Scarantino and Nielsen's accounts. The main question unanswered by Scarantino and Nielsen was the following: Why do we sometimes displace a thwarted action tendency *radically*—onto the closest thing at hand—and sometimes *symbolically*—taking one specific object to stand for another? The answer by my account is the following: We displace symbolically when we desire to do X with object A, and there is an object B in our surroundings that improves the imagining of doing X with A. In these cases, B will stand for A in the symbolic action in virtue of its *symbolic meaning*—the set of associations it triggers, which enrich the imaginative project the subject is immersed in. I will specify the relation between symbolic representation and symbolic meaning in the next subsection.

Regarding Goldie's account, there was one main unanswered question in explaining intrinsic symbolic actions. If we accept that desiring to X sometimes invites us to imagine doing X, why do we sometimes restrict ourselves to imagining "in the mind's eye" and other times engage in active imaginings? The answer is the following: The imagining turns into active pretense because there is a thwarted tendency to execute, and we displace it onto the symbolic action. To generalize:

- **Symbolically displaced imagining's account:** For any intrinsic symbolic action, the agent did S because she was under emotion E, which has X as its goal doing X was not possible, he imagined doing X, doing S improved the imagining of doing X while allowing for the displacement of the emotional action tendencies.

2.6.2. S*ymbolic* meaning and the conditions for symbolic representation

Hell is where nothing connects with nothing

T.S Eliot's introduction to Dante's inferno

In the previous sections (see 2.2) I have argued, contra Goldie and Scarantino and Nielsen, that, in symbolic actions, the relationship between the object at the center of the action and the object it stands for is not arbitrary nor can it be stipulated at will. This central feature distinguishes intrinsic symbolic actions from pretense since in the latter we can stipulate at will which object stands for another. Regarding symbolic actions, the following question remains unanswered: What are the conditions object B must meet to represent object A *symbolically*? Before answering this question, it would be useful to introduce the notion of *symbolic meaning.*

In colloquial language, we frequently say that "things" have *symbolic meaning* to us. "This house *means* a lot to me", someone may say. But what is it for an object to have *symbolic meaning*? The phenomenon by which objects acquire meaning for us seems quite existentially important and present in everyday expressions. However, to my knowledge, there is no attempt in the analytic tradition to give an account of this kind of meaning. It is an interesting case since *symbolic meaning* is neither intrinsic (as in the case of concepts), natural (as in the case of fire/smoke), nor conventional (as in the case of language)—to name a few. Rather, it seems somehow grounded in personal history and expressed in the triggering of memories and emotions. Making sense of this folk psychological expression would require a more extensive reflection. However, an initial characterization is the following: The symbolic meaning of an object is the set of associations, memories, and affect it triggers. Objects, in these pictures, are not just objects, but also catalysts for mental states, different for every person in virtue of the personal history with the object.[27] For this reason, we sometimes feel sad about losing an object, no matter its monetary value. The widower in [1] would be extremely sad about the loss of his dead wife's dress beyond its economic value and usefulness. The object, by triggering a specific set of memories, emotions, and associations provided a specific

[27] Cassirer seems to point to this phenomenon in a more obscure way when saying that, in confronting reality, the man is constantly conversing with himself: "No longer can man confront reality immediately; he cannot see it, as it were, face to face. Physical reality seems to recede in proportion as man's symbolic activity advances. Instead of dealing with the things themselves man is in a sense constantly conversing with himself."

mental coordinate. Once the object is gone, it might not be that easy to internally trigger the exact set of mental states (memories, emotions, images, and so on) the object used to trigger. The notion of *mental coordination* will be clearer with an example. I have my grandfather's shotgun and pen as keepsakes. Both objects remind me of my grandfather, however, they do so in very different ways. The shotgun brings to mind not only my grandfather but also his love of hunting, dead animals, and negative emotions. His pen brings to mind my grandfather, but also his stories, his handwriting, and his positive emotions. The objects have different *symbolic meanings:* They trigger a specific set of memories, associations, and emotions. Because of this, each object provides a specific coordinate from which to remember my grandfather. Once objects acquire symbolic meaning, understood in this sense, they are no longer reduced to their physical properties or usefulness. We also value them as the triggers for a specific set of mental states. We like to surround ourselves with objects that have a symbolic value for us because they trigger certain thoughts and emotions. For these reasons, if we do not know the symbolic meaning an object has for a person, it might be difficult to regard her action as intelligible, even in light of her reasons for the action.

Going back to our initial question: What are the conditions object B must meet to represent object A *symbolically*? Object B can symbolically represent object A in virtue of its *symbolic meaning*, which is the set of associations (e.g., memories, images, and emotions) it triggers in a specific agent. If object B triggers a set of associations that concerns object A and its relevant aspects of it, object B will be a suitable candidate for displacing action tendencies generated by object A. Because of its *symbolic meaning,* an object can serve as a substitute for the absent object in the symbolic action. In symbolic actions, the presence of the symbolic object B increases the vividness of the imagining of doing X with A and invites us to displace the action tendencies of the emotion towards it (doing S with B). For example, the desire to scratch out the eyes in the picture emerges because the picture—by providing a vivid image of the rival—feeds and intensifies the imagining of scratching the eyes of the person in the picture. Acting on these behavioral propensities promotes emotional regulation by allowing one to simultaneously imagine doing X with A and doing S with B (when B is a symbol for A).

It can be argued that for the relation of symbolic representation to exist, sometimes only resemblance is needed. This seems to be the case in [2] and [3]. In [2] the mere image of a person triggers the same set of associations and emotions seeing the person might trigger (to a lesser degree). The picture has symbolic meaning in a very poor sense. However, in a vast majority of cases, resemblance will not be enough for an object to symbolically represent another. The symbolic meaning of an object can change drastically with a small change in its properties and, with it, the symbolic actions it might trigger. Consider the following example: In a strict sense, a flag is just a colorful cloth. But a flag has also, for some people, symbolic meaning (emotions, memories, and ideas associated with it). The associations a symbol triggers can be immensely specific. In the case of the flag, a mere color change can frustrate the possibility of the flag standing for a specific set of ideals. In 1990, the artist David Hammons created an African American flag, just like the American flag, but with colors more typical of the African continent: red, green, and brown (see Picture 1). This flag is clearly very similar to the American flag. However, a color change can modify the set of associations and emotions triggered by it. Imagine the case of the patriot who, unironically kisses his flag out of happiness when his preferred political candidate wins. For the sake of the example, imagine that the patriotic subject is racist. Probably, regardless of its similarity with the American flag, Hamon's flag will not displace affection towards it, in his case. Although the resemblance between the absent object and the object of the symbolic action is sometimes enough, this is not the golden rule for symbolic displacement. This seems to depend on a whole set of associations triggered by objects—their symbolic meaning, which will vary for each object and person.

Picture 1: David Hammons, 1990.
African American Flag. MOMA.

2.6.3. What kind of satisfaction do we obtain in performing those actions?

Scarantino and Nielsen (2016) claim that what distinguishes *radically* from *symbolically* displaced emotional actions is that the latter provides a symbolic satisfaction of the relational goal of the emotion. However, Scarantino and Nielsen do not specify in what respect symbolic satisfaction is different from mere satisfaction. I share with them the intuition that the satisfaction of venting an emotion on the most proximal object differs from the one obtained when symbolically displacing the emotional action. In my account—simultaneously imaging performing an action and acting on the tendencies prompted by the emotion—can specify the notion of symbolic satisfaction a bit further. The subject obtains *satisfaction* because the emotional action tendencies are released. The satisfaction is *symbolic* because the person simultaneously imagines doing the thwarted action and the object of the action stands in a *symbolic* relationship with the object causing the emotion. The presence of the symbolic object contributes to the action by adding vividness and content to the simultaneous imagining—something that an object randomly chosen would not achieve. Doing S with B provides a symbolic satisfaction of the goal of doing X with A in virtue of the symbolic representative relationship between B and A. That is, the satisfaction is symbolic because the action tendency is displaced towards an object whose symbolic meaning (the set of associations it triggers) appeals to A.

One further question about symbolic satisfaction (the satisfaction we obtain when displacing an emotional action tendency from object A to an object representing it) is whether it can go unnoticed by the subject. It is crucial to note that if we accept that humans redirect emotional responses, we have no reason to claim that their redirection is restricted to inanimate objects. These redirected responses are more evident in the case of inanimate objects because the acts of redirection appear to be arational (Why hit a cushion? Why caress a dress?). However, displaced actions on other people would have a rational appearance since we can always come up with a candidate explanation. Although the legitimacy of these interpretations has not been examined in detail, the phenomenon is sketched by De Sousa (2017). In this example, the emotion toward a person (the mother) is displaced towards another (the wife) in a symbolic way:

> "if I am experiencing an emotion that seems altogether inappropriate to this occasion, I will naturally confabulate an explanation for it. A neurotic who is unreasonably angry with his wife because he unconsciously identifies her with his mother will not rest content with having no reason for his anger. Instead, he will make one up. Second, the reason he makes up will typically be one that is socially approved" (2017, p.31).

This brings the possibility of displacement being unconscious to the table. Can the symbolical displacement and the symbolical satisfaction obtained when performing an action be unknown by the subject? Can the trigger of emotion and the object or person symbolizing it be unconsciously identified as in De Sousa's example? These kinds of attributions are at play in psychoanalytical explanations of behavior, but their legitimacy and their cognitive ground have not been explored in analytic philosophy. Further investigation in this respect will help elucidate the epistemological legitimacy of such attributions, in which the subject is unaware of the displacement of an action tendency.

2.7. Conclusions

In this chapter, I have proposed an account of intrinsic symbolic actions which combines Goldie's (2010) and Scarantino and Nielsen's (2016) accounts. Previously, I have shown that these actions cannot be rendered intelligible by appealing to a belief-desire pair held by the agent nor by mentioning her emotional state. In my account, symbolic actions have a teleology but are caused non-rationally. They are a way to deal with the frustration concerning the impossibility of doing a certain action in the grip of an emotion. When performing symbolic actions, two phenomena occur synchronically. First, thwarted action tendencies are displaced in a non-arbitrary way and released. Second, the subject imagines herself performing the frustrating action while displacing these action tendencies. An account for the conditions object B must meet to be able to represent object A has been provided. The release of these action tendencies on an object which stands in a symbolic relationship with the object that causes the emotion provides *sui generis*, *symbolic* satisfaction. In this respect, symbolic actions are optimized imaginings, since besides imagining the achievement of a goal, they allow for a part of such goal to be consumed (the displaced action tendency). Further investigation of the notion of symbolic satisfaction and the possibility of displacement being unconscious will be useful in evaluating the legitimacy and cognitive ground for, for example, psychoanalytical practice.

Part II
Boundaries of imagination

Chapter 3

Placing delusion in a fragmented and psychofunctional system of belief

Abstract: The functional profile of delusions has motivated a philosophical debate in recent years on the status of delusions. The focus has been on whether these should be regarded as beliefs, as they are in psychiatry and psychology. The doxastic conception of delusion has been under criticism and alternative accounts have emerged. Such is the case of Currie and colleagues' non-doxastic metacognitive account, according to which delusions are imaginings misidentified as beliefs by the delusional subject (Currie 2000, Currie & Jureidini 2001, Currie and Ravenscroft 2002). My aim in the present chapter is twofold. First, I raise a critique in response to Currie and colleagues' (2000, 2001, 2002) claim that imaginings describe the functional profile of delusions better. More specifically, I show that Curie and colleagues' (2000, 2001, 2002) metacognitive account is not well equipped to explain delusional incorrigibility. Second, I attempt to model delusions in a fragmented model of belief. In doing so, I raise criticisms of Davies and Egan's (2013) doxastic and Bayesian account of delusions in their fragmented system. I argue that Davies and Egan's commitment to Bayesian laws of belief formation and revision hinders its ability to explain this phenomenon. I then sketch my positive proposal: to model delusions in a fragmented, albeit non-Bayesian, system of belief (Mandelbaum and

Bendaña 2020, Mandelbaum 2019). I conclude that Mandelbaum and Bendaña's model can account for the functional profile of delusions from a doxastic perspective better, since it allows for the influence of motivational factors in belief acquisition and updating.

3.1. Are delusions beliefs?

Patients who suffer from delusions claim to believe the states of affairs that do not align with reality. They can assert with conviction being Jesus Christ (Young 2000), being dead (Young et al. 1994), or that their relatives have been replaced by look-alike impostors (Coltheart 2007). Patients who suffer from delusions sustain their statements in a firm, recalcitrant manner, despite available evidence against the delusion. Fortunately, the firmness of their verbal reports does not match their behavior: Delusions generally have weaker effects on action than expected (Sass 1994; Bayne & Pacherie 2005; Bortolotti 2010, Stone and Young 1997).

Doxasticism about delusion is the theoretical stance according to which delusions are *believed* by the subject reporting them. Doxasticism is the dominant psychiatric and psychological view, reflected by the DSM-V definition (Schizophrenia Spectrum and Other Psychotic Disorders, 2013):

> "Delusions are fixed beliefs that are not amenable to change in light of conflictingevidence. […] Delusions are deemed bizarre if they are clearly implausible and not understandable to same-culture peers and do not derive from ordinary life experiences. […] The distinction between a delusion and a strongly held idea is sometimes difficult to make and depends in part on the degree of conviction withwhich the belief is held despite clear or reasonable contradictory evidence regarding its veracity" (American Psychiatric Association, 2013, p. 87).

Although doxasticism is the traditional stance in the psychiatric literature, it has been a point of contention in the philosophical debate, where it has met with several objections. The reason for this debate is that delusions do not appear to have the expected functional profile of beliefs. More specifically, they do not seem to satisfy the following rationality constraints: They do not integrate with the rest of the patient's beliefs, they are unresponsive to evidence, and 3they do not guide action across contexts (Dub 2017, Bortolotti 2018). For these reasons, the doxastic conception of delusions is still under debate, and several doxastic and non-doxastic accounts have emerged addressing these objections (Bongiorno 2021; Smithies, Lenon & Samuels 2022).

Historically, doxastic accounts have focused on two main questions: What is the cause of the delusional belief and why does it persist over time despite encountering counterevidence. Two-factor accounts propose two factors in response to these questions (Coltheart, 2007; Coltheart et al., 2011; McKay, 2012). A third issue has been added in light of the criticisms doxasticism has faced: Although delusions may sometimes lead to action, often times they fail to exhibit the full behavioral repertoire that would be predicted if delusional subjects were fully committed to the content of their beliefs. For instance, patients with Capgras' syndrome claim that a relative has been replaced by a replicant, and yet accept to peacefully live with the replicant and/or fail to inquire about the fate of the missing one. To fully account for delusion, an explanation needs to be provided regarding patients' failure to act on their professed delusions, a phenomenon known in the literature as "double bookkeeping". To date, there is no consensus on a particular doxastic account meeting these requirements in a satisfactory way. My main aim in this chapter is show that Bendaña and Mandelbaum's fragmented and psychofunctional system of belief, if correct, offers a good framework for modelling delusions and explaining their functional profile qua beliefs.

To better understand the phenomenon described, in the next section, I characterize delusions in the context of the Capgras Delusion. I then proceed to formulate and assess Currie and co-authors' imagination account to conclude that it cannot explain the persistence of delusions despite the availability of overwhelming counterevidence. In section 4, I evaluate the potential of two fragmented systems of belief to model delusions. I point to the limitations of Davies and Egan's (2013) Bayesian and fragmented account and conclude that doxasticism about delusions can be better modeled in Bendaña and Mandelbaum's (2021) fragmented and psychofunctional system of belief.

3.2. The Capgras Delusion: Circumscription, resistance to evidence, and double bookkeeping

The Capgras delusion is the pathological condition in which a patient reports that a close friend or relative—frequently their spouse—has been replaced by a look-alike impostor. This delusion is considered the Rosetta Stone in the study of delusions given its circumscription, monothematicity, and well-established neuropsychological origin (Coltheart, Menzies, Sutton 2010: 267). In monothematic delusions, the patient exhibits either one or a set of delusional beliefs related to a *single theme*. In contrast, in polythematic delusions, subjects manifest delusional beliefs about a variety of—often unrelated—topics. Because of its paradigmatic features, research has focused on the Capgras delusion, assuming that the conclusions obtained will generalize to other monothematic delusions of neuropsychological origin. Following this reasoning, I will only refer to this delusion.

3.2.1. On the origin of delusional belief: An abductive inference from an anomalous experience of neuropsychological origin

It is well established that Capgras delusion is generated in response to an anomalous neuropsychological experience (Stone and Young 1997; Coltheart and Davies 2022). In non-delusional subjects, seeing and recognizing a familiar face causes a response in the person's autonomic nervous system, which factors in affective terms in the overall phenomenology. Ellis and Young (1990) were proposed that the nature of this abnormal experience in patients with Capgras delusion was a lack of this autonomic, affective response when seeing a familiar face. Several experiments measuring skin conductance reactions to faces (Brighetti et al., 2007; Ellis et al. 1997, Hirstein & Ramachandran 1997) confirmed this hypothesis in the following years. There is now a consensus that the face recognition system is disconnected from the autonomic nervous system involved in affective processing in Capgras patients. Although the systems themselves remain intact, the connection between them is broken (Coltheart at al. 2011, Coltheart and Davies 2022). As a result, patients recognize the identity of the face—its physical features and the fact that they match a stored representation of a relative—but lack the affective response that usually accompanies the perception of familiar faces. This generates a mismatch between the semantic and the affective information in patients with this kind of delusion. On the one hand, they physically recognize the person but, on the other, they do not feel the expected familiarity when seeing them. The lack of this autonomic response is not explicitly noticed by the subject, given that we are not consciously aware of our autonomic responses. However, this absence can influence the conscious experience, contributing to a phenomenological feeling of strangeness: The person *looks like* their relative but does not *feel like* their relative. This paradoxical situation constitutes the abnormal experience, and it is taken to be the cause of the delusional belief. This belief results from a reasoning process that takes place to find a hypothesis for the abnormal experience:

> "When patients find themselves in such a conflict (that is, receiving some information which indicates that the face in front of them belongs to X, but not receiving confirmation of this), they may adopt some sort of rationalization strategy in which the individual before them is deemed to be an impostor, a dummy, a robot, or

> whatever extant technology may suggest" (Ellis & Young, 1990, p. 244)

A patient with Capgras syndrome that believes her husband is an impostor in disguise might follow this reasoning: "This man *looks like* my husband but *feels like* a stranger. The best hypothesis explaining this is that, although he has my husband's appearance, he is a stranger. Therefore, he must be an impostor or a robot".[28] This misidentification— namely, the delusional denial regarding the authenticity of a recognized relative—is construed as an interpretative illusion and not as a misperception of external stimuli. This interpretative illusion occurs when the subject makes an inference from the paradoxical experience she has when *recognizing* a well acquainted or highly familiar face and not *feeling* the expected affective response in the overall phenomenology of the experience.[29]

The acknowledgement of abnormal neuropsychological functioning explains the origin of the Capgras delusion. This is also the case for other monothematic delusions of similar origin—namely, a malfunction in the autonomic system. Such is the case of the Fregoli delusion, the inverse of Capgras. In Fregoli, the autonomic system is overresponsive to faces. Because of this, even faces of strangers produce autonomic responses, causing the patient to experience these faces as highly familiar. This abnormal experience is rationalized by the subject with the delusion that strangers around him are familiar persons in disguise (Langdon et al.

[28] There is a discrepancy in the literature on whether the delusion results from an endorsement of the abnormal experience (Bayne and Pacherie 2004; Pacherie et al. 2006) or as an explanation of it (Maher 1999; Stone and Young 1997). I will leave this issue aside given that origin of the delusion is not relevant to the debate between the examined accounts.

[29]Some think that the transition from the abnormal experience to the delusional belief is a rational one (Coltheart et al. 2010, Davies and Egan 2013). According to this view, the abnormal experience constitutes enough evidence to support the delusional hypothesis. In contrast, others claim that the adoption of the delusional belief does not follow normal reasoning processes and some further impairment or bias is needed to explaining it (McKay 2012, Parrot 2014). I will set this issue aside for this chapter.

2014).[30]

3.2.2. Incorrigibility and inconsequentiality in Capgras

A second intriguing issue with respect to Capgras concerns the persistence of the delusion over time, despite the availability of overwhelming evidence against it (Coltheart 2007, Pacherie 2009). In Capgras, once the delusional belief is adopted, the patient receives evidence that should led her to abandon the delusional belief—hereinafter, the impostor hypothesis. Doctors, friends, and other family members will give her testimonial evidence against this hypothesis. They would provide information about the neuropsychological origin of the delusion, for example. Furthermore, the Capgras patient will find new firsthand evidence against the content of the delusion. In talking to the alleged impostor, she will be able to check that his personality and memories are the same as those of her relative. This should invite her to abandon the impostor hypothesis (namely, "this man looks like my husband but is not my husband") and embrace the relative hypothesis ("this man looks like my husband and is my husband"). However, patients seem to ignore this evidence and the delusional belief shows incorrigibility (Poupart et al. 2021: 3). Even when the patient realizes that some of her beliefs go against the delusion (e.g., the belief that it is impossible that a stranger knows her history perfectly if he is not her husband), there is no belief revision to restore consistency. In contrast, patients with Capgras syndrome show little interest and even resistance to resolving contradictions between their delusion and other background beliefs (Bayne and Pacherie 2005).

When considering behavior in the Capgras delusion, the same inconsistency is observed. For example, one would expect that patients

[30] Another delusion explained by the malfunctioning in the autonomic system is the Cotard delusion. In this case, the patient claims to be dead due to autonomic non-responsiveness to almost any stimulus.

would refuse to live with the alleged impostor and show resistance—violently, if needed—to living with a stranger in disguise. More importantly, if she liked her husband, she should desperately look for him, since he has disappeared and been supplanted by an impostor. However, often none of this happens. Patients with Capgras syndrome do not display the full set of actions congruent with their delusion (Alexander et al. 1979, Bayne and Pacherie 2004, Poupart et at. 2021). Patients often show no reluctance to living with a replicant—or a group of them—and, despite thinking that they are living with mere replicants, they do not search for their "real" relatives.[31]

Let us now see how Curie and colleagues' (2000, 2001, 2002) metacognitive account accommodates for the circumscription, incorrigibility, and behavioral correlates of delusions.

3.3. Delusions, imaginings, and responsiveness to evidence

The functional differences between delusions and *bona fide* beliefs have motivated alternative accounts on the nature of delusions. These accounts deny that subjects believe the content of their delusions. The argument main against doxasticism tends to have the following form:

P1) Beliefs have certain characteristics: They are formed and revised based on evidence and they guide action in the relevant circumstances.

P2) Delusions do not exhibit these characteristics.

C) Therefore, delusions are not beliefs.

[31] This pattern can also be seen in delusions of grandeur: The patient who believes he is Jesus Christ does not treat the doctor as a disciple, and the patient who claims his is Napoleon does not speak to crowds as if they were his troops (Young 2000). Likewise, patients who claim to be a dog do not show a disposition to bark (Bleuer, 1924). Even though in verbal reports they firmly support holding the delusional belief, they do not exhibit the expected behaviors.

This criticism of doxasticism is usually followed by a positive non-doxastic account. This is the case of Currie and colleagues' metacognitive account of delusions. According to them, delusions are not believed but rather merely imagined and misidentified by patients as beliefs (Currie 2000, Currie and Jureidini 2001, Currie and Ravnescroft 2002). Contrary to beliefs, they claim, imaginings perfectly match the functional profile exhibited by delusions. In this section, I will question the fit Currie and colleagues (2000, 2001, 2002) suggest between delusions and imaginings.

3.3.1. Currie and colleagues' metacognitive account

According to Currie and collaborators (2000, 2001, 2002), the victim of the Capgras delusion, who claims that his wife has been replaced by an impostor, does not erroneously believe that his wife has been replaced by an impostor; he merely imagines that his wife has been replaced, and erroneously identifies this imagining as a belief. In other words, the patient suffering from delusions has a meta-belief—he believes that he believes in the content of the delusion—but lacks the first-order belief in the delusion, which is merely imagined. The verbal behavior of the patient suffering from delusions is explained by a metacognitive mistake regarding the nature of the delusion. Although the patient lacks the first-order belief in the delusion (e.g., she does not believe that her husband has been replaced by a look-alike impostor), she has a higher order belief in the content of their delusion, which explains her verbal reports (she believes that she believes that her husband has been replaced by a look-alike impostor). The lack of a first-order belief in the delusion explains why delusions are circumscribed and why patients fail to act on the content of their delusions: They do not believe this content but rather imagine it.

Currie and colleagues' account has already been criticized for the idealized picture of beliefs it depicts (Bortolotti 2009, 2011) and the demanding rationality constraints on beliefs (Smithies, Lenon, Samuels

2022). Others have also emphasized that delusions are not as inert as they have been portrayed in leading to action (Poupart et al. 2021). I will address these concerns at the end of this section. My first aim is to question Currie's positive thesis based on different criticisms.

3.3.2. Epistemic uses of imagination and resistance to evidence

According to Curie and Jureidini (2001:160), the Capgras delusion emerges as a hypothesis in response to the odd experience that a relative's face *looks like* a relative but does not *feel like* the relative. In attempting to account for this experience, the subject imagines a hypothesis about what could account for the experience:

> "The idea that one's close relatives have been replaced by aliens of similar appearance *accounts for* one's peculiarly unemotional response to their presence" (Curie and Jureidini 2002: 159, emphasis added).

Currie and colleagues claim that the imagining status of delusions explains why subjects do not try to resolve tensions between the delusion and their other inconsistent beliefs (Currie and Ravenscroft 2002: 178). There is no requirement for the content of one's imaginings to be consistent with the rest of one's beliefs. For instance, my imagining that I have already received my doctorate degree will not start a process of belief revision to render the imagining consistent with the belief that I have to write a dissertation. The imagining and the belief of my doctoral student status can coexist. Furthermore, we do not tend to act on what we imagine. My imagining that I am already a doctor will not make me stop writing my dissertation. So, the lack of delusion-generated activity observed in patients is also explained if imaginings are delusions. Crucially, Curie and Jureidini (2001) claim that the unresponsiveness to evidence exhibited in delusions can also be accounted for by delusions being imaginings:

> "Finally, imaginings are not apt to be revised in the light of evidence;

> the whole point of imagining is to *enable us to engage with scenarios that we know to be non-actual*. Thus, imaginings seem just the right things to play the role of delusional thoughts; it is of their natures to coexist with the beliefs they contradict, to leave their possessors unwilling to resolve the inconsistency, and to be immune to conventional appeals to reason and evidence." (Currie and Jureidini 2001: 160, emphasis added)

In explaining the origin of the delusion, Currie and Jureidini's argue that the imagining emerges as an explanatory hypothesis about the *real* cause of a *real* odd experience. When accounting for the delusion's resistance to change on the face of counterevidence, they emphasize that the point of imagining is to *enable us to engage with scenarios that we know to be non-actual*. In Currie and Jureidini' account (2001), the imagining that constitutes the delusion seems to have two opposite natures. On the one hand, it emerges as a potential cause for an *actual* experience and therefore, it is to be expected that the subject will constrain this imagining with knowledge on how the world is. On the other hand, it is not responsive to evidence because, as an imagining, it concerns nonactual scenarios.

If the delusion emerged as a hypothesis of what causes the *abnormal experience*—as Currie and Jureidini claim—, it should be able to integrate and respond to new evidence. We frequently engage in imaginings that aim at grasping the causes of our experiences. Because of the use we make of them, these imaginings take new evidence into account in order to reassess the actual state of affairs. Take for instance the following case. David, an unfriendly person, arrives to work in a particularly good mood. Seeing him in this mood is a very rare experience for his subordinate Joana, who is used to seeing him in a lousy mood. To explain David's good mood, Joana entertains potential explanations. For instance, she imagines that David has just quit his job and is just picking up his belongings; this is why he is happy. Since this is a merely imagined hypothesis, Joana's belief that David is still her boss will not be affected. Moreover, since it is only an imagination, Joana will not behave as if it were true. For example, she will not tell her colleagues that David has quit his job. However, since the *raison d'être* of the imagining is to account for a certain observation, the imagining will remain permeable to evidence. If Joana learns, for example, that David is scheduling a meeting for next week, the imagining will be modified, and another imagined hypothesis will replace the previous one. Because the imagining aims at grasping the cause of a *current* event or experience, it is in the interest of the imaginer to anchor the imagining with as much real-world information as possible.

Given the previous argument, Currie and colleagues' positive proposal cannot account simultaneously for the origin of the delusion (an imagined hypothesis about the *actual* cause of an *actual* abnormal experience), and its resistance to change in light of counterevidence. Characterizing the delusion as a hypothesis of the cause of an abnormal experience conceptualizes delusion as an epistemic use of imagination (Badura and Kind 2021). Such imaginative projects are characterized as obeying or at least *trying to obey* certain constraints (Kind 2016: 151). More specifically, they are governed by the Reality Constraint, which states that *the world is imagined as it is*. When we make epistemic uses of imagination to discover the cause of a certain experience or observation, we constrain the imagining with as much knowledge about the real world as possible (as in Joana's imagining). By the same logic, Capgras delusion should be open to change based on new evidence. Even if the imagining that constitutes the delusion remains circumscribed, new evidence should refine it. However, as we have seen, the Capgras delusion remains unchanged in the face of counterevidence.

Leaving aside the plausibility of Currie and colleagues' criticism on doxasticisim, there are counterarguments to their positive account. If delusions were the kind of imaginings they claim there are, they would be changed in light of new evidence. Consequently, Currie and colleagues need to explain why the imagining that constitutes the delusion is fixated, just like the doxastic account needs to explain the persistence of the belief in light of counterevidence.

I will now turn to other criticisms that nondoxastic accounts such as Curie and colleagues' have faced, mainly given their idealization of the functional profile of beliefs.

3.3.3. Delusions and context sensitivity

Currie and coauthor's account has focused on the fact that delusions are somewhat realistic, non-idealized everyday beliefs (Bortolotti, 2010). Bortolotti claims that we do not need to give up the doxastic account of delusions when faced with evidence of patients who do not act on their delusions, since this also affects ordinary beliefs (Bortolotti 2011). It has been claimed that the behavioral inertness invoked by nondoxastic accounts is not systematically observed in patients with Capgras syndrome, who sometimes have safety-seeking or violent behaviors (de Pauw and Szulecka 1988). Although non-violent patients constitute the majority and, in most cases, they seem to live peaceably with the alleged imposter (Förstl et al 1991, Pandis et al. 2018), the evidence regarding violent behaviors is disputed in the literature (Poupart et al. 2021). What is undisputed is that the average patient with Capgras delusions does not show the full range of expected behaviors across contexts. Considering this, the behavioral profile of the Capgras delusion is better explained as being intermittent or context sensitive rather than inert.

There are alternatives to doxastic accounts explaining why the delusional belief does not guide action across contexts. One of them is that delusions are not well integrated with other beliefs. It is possible to excuse the behavioral dispositions by positing that the delusional belief is exceptionally encapsulated (Bayne and Pacherie 2005). Another alternative is to defend that our system of beliefs is by nature fragmented. Because of this, beliefs residing in different fragments can guide action differently depending on the context (Davies and Egan, 2013, Bendaña and Mandelbaum 2021). In the next section, I will address the explanatory potential of fragmented accounts of belief storage (Davies and Egan 2013, Bendaña and Mandelbaum 2021) for the functional profile of delusions.

3.4. Delusions in a fragmented system of belief

The widely accepted dogma in philosophy is that our belief system is unified in a web of consistent interconnected beliefs. According to this unified model, action and reasoning are guided by the entirety of interconnected beliefs that belong to a single network. Consequently, action should be consistent across contexts. Therefore, to accommodate for action-inconsistency, the mind must store inconsistent beliefs in different parts of a fragmented system. In these kinds of fragmented systems of belief (Egan 2008, Egan 2021, Bendaña and Mandelbaum 2021), the subject's total set of beliefs is compartmentalized into various subsystems (fragments). Because of this compartmentalization, the human mind does not need to be a logically consistent and deductively closed system. Inconsistent beliefs can be sustained if they are stored in different fragments. If a person holds the belief that P in a fragment of their mind and, in another fragment the belief that ¬P, it can be expected that, depending on which fragment is activated in a context, the subject will act one way or another.

The fragmented model, therefore, is *a priori* well equipped to account for inconsistent behavior across contexts and the circumscription of belief in Capgras.[32] The fact that delusional patients are doxastically field dependent (Bayne & Pacherie 2005) fits very well with the idea that our

[32] It is not my aim to make a comparison between two models of belief storage, nor to claim that one is better than the other overall. My aim is simply to see the advantages of the fragmented model in accounting for the Capgras delusion. Undoubtedly, the unified model has independent advantages over the fragmented model. For instance, it can account better for how people reason about many varied and unrelated topics at a time.

belief system is fragmented.[33] If beliefs are stored in different fragments and there can be inter-fragment inconsistency, we can appeal to this to explain the behavioral inconsistency seen in patients who display Capgras delusions. For these reasons, it is worth exploring whether placing delusions in a fragmented system could explain their functional profile.

Fragmented accounts are also well equipped to explain reported cases in which even the verbal report of the delusional belief seemed to fluctuate across contexts:

> "(…) 34-year-old son of the family who sometimes expressed the belief that his mother, father, and sister had been replaced by impostors, but at other times correctly identified them as genuine family members. He would sometimes go into their bedrooms at night and shine a torch on the sleeping person's face in order to determine whether it was the impostor or the genuine family member who was there. This man's beliefs about his family members fluctuated between being correct and being Capgras-delusional" (Coltheart, 2007).

In cases like this, the delusional belief would be responsible for the subject's verbal behavior and some other dispositions in specific contexts, and the opposite belief would be responsible for other behaviors, such as agreeing to live peacefully with the alleged impostor, or not engaging in a search for his disappeared relatives.

In the following, two recent fragmented systems of belief will be considered. The first one has already been applied by its authors to give an account of delusions. Davies and Egan (2013) combine a fragmented model of belief storage with a Bayesian approach of belief acquisition and updating in delusions. Davies and Egan are committed to high standards of

[33] The motivations for a fragmented systems of belief storage are of course independent from the topic that concerns us here (to get a glimpse of the motivations for this picture of the mind see: Borgoni, Kindermann, and Onofri 2021.

rationality, adhering to Bayesian views. The other case (Bendaña and Mandelbaum, 2021) combines fragmentation with a non-Bayesian approach. Buongiorno (2022) has recently modeled delusions this way, but my account will differ from his. In my view, the central benefit in modelling the Capgras delusion within this framework is that it allows for psychological motivations to have a role in belief revision and updating, which can explain delusion's circumscription and resistance to counterevidence better.

3.4.1. Delusion in Davies and Egan fragmented and Bayesian system of belief

Davies and Egan (2013) have recently offered a doxastic account of delusions in a Bayesian and fragmented system of belief. In their fragmented model there is no need for consistency among different fragments: A belief and its opposite can be held by a subject if they are stored in different fragments. At the same time, a certain belief and its contrary can guide action in a context dependent way, according to which fragment is activated:

> "Actual belief systems are fragmented or compartmentalized. Individual fragments are consistent and coherent, but fragments are not consistent or coherentwith each other and different fragments guide action in different contexts. We hold inconsistent beliefs and act in some contexts based on the belief that P and in other contexts on the basis of the belief that not-P" (Davies and Egan 2013: 705)

This kind of model allows for the possibility of action being guided by different beliefs in accounting for the inconsistency between verbal and non-verbal behavior in patients with Capgras delusions. Additionally, Davies and Egan's account needs to explain why the delusional belief is adopted and why it is maintained on the face of counterevidence. It is to be expected that in certain situations—in interviews with doctors, for example—the patient would be invited to activate the fragment containing the delusional belief ("This man—who looks like my husband—is not my husband") and the fragment containing the opposite belief ("The man—who looks like my husband—is my husband"). The coactivation of fragments should render the beliefs in those fragments consistent. This means that it should eradicate either the delusional belief or of its opposite. Davies and Egan account finds some difficulties explaining why this is not the case in Capgras.

Concerning the adoption of the delusional belief in Capgras, Davies and Egan claim that this belief is automatically formed and compartmentalized following the abnormal experience. They characterize the belief as a prepotent doxastic response to the experience of strangeness when confronted with a familiar face (Davies and Egan 2013: p. 714). However, they need to explain why the delusional subject does not reevaluate this belief when contrasting it with background beliefs or evidence presented by doctors and relatives. This is especially relevant given that they are committed to the Bayesian model of belief evaluation and updating. The Bayesian framework is thought to let us determine the most rational update in which a subject should adjust his or her beliefs when facing new evidence, no matter how rare or abnormal this evidence is. The Bayesian model understands beliefs as subjective probabilities or levels of credence assigned to possible hypotheses. The limits of the probability space (one and zero) correspond to all or none beliefs. When faced with new evidence, the prior level of credence assigned to the hypothesis should be adjusted. This means that, before the abnormal Capgras experience, the subject assigns a certain probability to the following hypothesis.

> H1: This man, who looks like my husband, is my
> husband.
>
> H2: The man, who looks like my husband, is not

my husbsand (*impostor hypothesis*).

It is to be expected that before the abnormal experience, the subject assigns a probability of one to H1 and zero to H2; these are the prior levels of credence. As previously mentioned, Davies and Egan consider that the delusional belief is formed as a prepotent doxastic response that does need to follow this Bayesian process. Namely, the exposure to abnormal experience (the experience of seeing a relative in patients with Capgras) can drastically change the levels of credence that we assign to this hypothesis, leading the subject to believe H2. However, since they endorse a fragmented picture of belief, they claim that this belief is immediately compartmentalized. Because of this, the prior probabilitiesof H1 and H2 remain intact in another fragment, and the patient still uses them at a later stage. According to their fragmented model, preexisting beliefs (prior levels of credence in H1 and H2) are retained and available to the subject in posterior stages of reevaluation of the delusional belief.

Combining fragmentation with the Bayesian approach requires that patients engage in multiple assignments of credence (different subjective probabilities concerning H1 and H2), stored in different fragments. After the abnormal experience and the formation of the delusional belief (which can be defined as a high level of credence in H2), the subject has access to the prior probability of H2, which was very low and thus should invite him to abandon H2. In other words, the subject should be in principle able to coactivate the fragment containing the delusional belief (high level of credence in H2) and the fragment containing beliefs previous to the abnormal experience (low level of credence in H2). Davies and Egan's (2013) fragmented system is committed to intra-fragment consistency—i.e., consistency among beliefs stored in the same fragment. On the other hand, it is also committed to merging both fragments when they are coactivated. For this reason, it is to be expected that if the prior low level of credence in H2 and the delusional belief are coactivated, the first would undermine the latter, inviting the patient to abandon the delusional belief.

Furthermore, when verbally stating the delusion, the patient would be invited (by doctors and other acquaintances) to simultaneously access the delusional belief and fragments that contain evidence that contradicts it. In this case, the belief should be reevaluated and eradicated. But none of this happens: The delusional belief is firmly defended. Regarding persistency, Davies and Egan have yet to explain delusions' resistance to revision and their persistence over time along with the opposite belief, especially considering the likelihood that the patient has coactivated fragments containing contradictory beliefs.

To explain why the coactivation of fragments in favor and against H2 does not cause belief consistency, Davies and Egan propose two possible abnormalities: Either the patient suffers cognitive impairments or fails in the compartmentalization of the belief—i.e., in his assignation of it to a fragment. According to this possibility, the belief would have been incorrectly compartmentalized in a way that makes it ubiquitous and, thus, fully integrated into the belief system. In turn, this would suppose the elimination of any other assigned levels of credence that would render H2 implausible (prior levels of credence). In the following, I will show that none of these options is satisfactory in accounting for the persistence of the delusion.

a) The postulation of unjustified cognitive impairments in patients with Capgras delusions

The maintenance of delusional beliefs in the face of counterevidence invited Davies and Egan to hypothesize the presence of cognitive impairment. This cognitive impairment is meant to account for failures in belief revision—namely, the fact that the coactivation of the fragment containing the delusional belief and the fragment containing the opposite belief does not eradicate the delusional belief (or its opposite). Davies and Egan suggest that the malfunctioning of the working memory system or executive functions affects the critical evaluation of the delusional belief.

However, positing this cognitive impairment is—in light of the evidence we have so far of Capgras—*ad hoc* and unjustified. First, we are dealing with a monothematic delusion, namely, one in which the patient's delusional beliefs concern only one topic. If the patient suffered from a general impairment in the working memory or executive systems, we would expect a whole range of effects beyond the domain of the delusion. But, as Pacherie indicates, this is not the case:

> "…the Capgras delusion, like other monothematic delusions, tends to be relatively circumscribed. In domains other than that of their delusions, the reasoning skills and cognitive behavior of Capgras patients appear, by and large, to be normal. What needs explaining is therefore not only why subjects fail to check their delusional belief appropriately, but also why the failure is localized." (Pacherie 2009: p. 119)

General cognitive functions are not impaired in delusional patients, who "usually maintain clear consciousness, with apparently intact cognitive functions" (Salvatore 2013: 2). There is a general consensus in the monothematic delusion literature that, despiteof the delusional belief, "the patient's cognition seems otherwise perfectly normal" (Coltheart et al. 2010). Furthermore, the delusion is considered to "[coexist] with a maintained contact with reality" (Poupart et al. 2021). Even Davies and Egan (2013) define monothematic delusions as "islands of delusions in a sea of apparent normality" (2013: 690). Therefore, explaining of the persistence of the delusional belief through a malfunction in the working memory system or executive functions is not justified by current evidence.

Let us now examine the second possibility the authors consider when explaining why the delusional belief is not corrected when faced with counterevidence.

b) Delusion and contextual action-guidance

Davies and Egan's (2013) other account for the lack of revision of the delusional belief concerns the possibility of a failure in its compartmentalization. They claim that when adopted, the delusional belief may have been fully integrated with all other beliefs. This means that it might have entered other fragments that contain beliefs inconsistent with the delusional belief, eradicating them. However, if this compartmentalization failure had happened, the limited effect of the delusional belief on behavior can no longer be explained. If the belief had been fully integrated within the belief network, we should expect it to have a more widespread effect on actions than it appears to. By appealing to the ubiquity of the belief in order to account for its persistency, Davies and Egan's (2013) account loses its explanatory power over the behavioral profile of delusions.

Overall, even though Davies and Egan's (2013) fragmented model can initially account for the subject's behavior by appealing to the compartmentalization of beliefs and context dependency, the model fails to account for the persistency of the delusional belief. Rather, it either postulates unmotivated cognitive impairments or explains the persistence of the delusion by appealing to a failure in compartmentalization that leads to fully integrating the belief. This last option takes away the explanatory potential from the model with respect to patient behavior.

3.4.2. Delusion in a fragmented and psychofunctional system of belief

After having critically assessed the explanatory potential of Davies and Egan's account, I will propose another fragmented system of belief, better suited to account for the Capgras delusion. Specifically, I will place the Capgras delusion within Bendaña and Mandelbaum's (2021) fragmented system of belief. Bendaña and Mandelbaum's system allows for inconsistency among beliefs stored in different fragments (interfragment inconsistency) but claims consistency among beliefs within a fragment (intrafragment consistency). That is, if two fragments are coactivated, they should be rendered consistent, ruling out contradictory beliefs. The main difference with Davies and Egan's system is that Bendaña and Mandelbaum's does not consider belief acquisition and updating to be strictly governed by Bayesian rules. In contrast, psychological principles have a central role, as Mandelbaum explicitly states:

> "The principles of belief acquisition and updating seem grounded in maintaining a psychological immune system rather than approximating a Bayesian processor" (Mandelbaum 2019: 1).

Several findings argue against the idea of the mind working as a perfect Bayesian processor (for a review see Mandelbaum 2019). Multiple evidence shows that, many times, we stubbornly adhere to our beliefs when faced with counterevidence in a way that is far from resembling a Bayesian processor. People, for instance, surprisingly increase their belief P after receiving information that not-P (Taber & Lodge 2006). The huge number of cases in which we appear to be updating our beliefs irrationally has invited some to consider the possibility that such outputs might not be errors in the processing, but rather part of its proper functioning (Quilty-Dunn & Mandelbaum 2017). This function is postulated to be part of a *psychological immune system*, according to which people will adjust their beliefs to avoid psychological discomfort. According to Bendaña and Mandelbaum (2021), this defensive system is constituted by core beliefs about the self, such as being a good person, being consistent, and being smart, among others. When these core beliefs are threatened, the psychological immune system will be activated "to ward off serious threats to one's sense of self" (Mandelbaum 2019:12).

One of the principles of Bendaña and Mandelbaum's fragmented model is that it allows for Representational Redundancy. This means that different tokens of a particular belief may be stored in different fragments. On the other hand, they postulate the principle of Multiple Resistance: The more redundantly represented a belief is, the more resistant to revision. That is, the more distributed and repeated the representation of a belief, the stronger. Bendaña and Mandelbaum claim that *core beliefs about the self* are ubiquitous. In other words, they are part of every fragment, making them permanently accessible. According to this, one's representation of the self (or self-concept) is extremely redundant; the self, in this account, is "the center of doxastic gravity" (2021: 90). Because of this, the belief that one is a good, smart, reliable person (and any other central trait of our self-concept) is accessible in any reasoning processes. The authors exemplify the phenomenon using a case of effort justification. In this example, a person joins the Marines. Surprisingly, after an unpleasant initiation ritual, this person does not dislike the Marines, but rather likes them more. According to Bendaña and Mandelbaum, the person's reasoning is the following (2021: 92):

> P1) I put a lot of effort into joining the Marines.
>
> P2) Only an idiot would put a lot of effort into joining the Marines without liking the Marines.
>
> P3) I am not an idiot.
>
> C) I must like the Marines. Thus, the opinion of the group is improved.

In such case, the process of belief change (the increased liking of the Marines) is caused by the desire to avoid believing that one is an idiot (because of the useless voluntary sacrifice) and the central, core belief that one is not an idiot. As we will explore in the next section, if one of the premises is ruled out in this kind of argument, the conclusion does not follow. Indeed, Bendaña and Mandelbaum report that participants with low self-esteem—i.e., those who are prone to accepting the label of idiot, against proposition P3—do not seem to follow this kind of reasoning. Therefore, they do not tend to show the normal effort justification effect. In the following section, I will explore the potential role of the *psychological immune system* in the persistence of the Capgras delusion in Bendaña and Mandelbaum's fragmented system of belief.

a) Fragmentation of belief and the safeguarding of the self-concept: adoption and persistence of the Capgras delusion

Motivational factors and self-deception have been proposed as accounts for several delusions (Bortolotti & Mameli 2012, Bayne and Fernandez 2010). As far as Capgras delusion is concerned, theorists have postulated defensive motives to account for its origin from the beginning. Initially the delusion was interpreted in psychodynamic terms. From a Freudian stance, the delusion was described as an attempt to veil incestuous desires for familiar members (Capgras & Carette, 1924). Once the relative was taken to be a stranger, the desire could be embraced without guilt. Reports of cases involving the Capgras delusion with animals—cats, parrots, and canaries, or even inanimate objects—argue against the psychodynamic interpretation (Abed & Fewtrell, 1990, Islam et al. 2015), and the Freudian hypothesis was soon abandoned. Since then, and given its well-studied neuropsychological origin, motivation has no longer played a role in accounts of the Capgras delusion (Mele 2006).[34]

However, I want to suggest that motivational factors could intervene, not in the origin, but in the maintenance of the delusion. The point I want to stress here is that we can account for the persistence of the

[34] Nevertheless, motivational accounts have been used to explain other delusions such as erotomania (Cléarambault's delusion). In this case, patients form the belief that someone of higher social status is secretly in love with them (Berrios & Kennedy, 2003; de Cléarambault, 1921/1942). Similarly, in persecutory delusions (Kinderman & Bentall, 1996) motivational factors—such as to maintain a positive self-image—have been used to explain the origin of the delusion.

delusional belief using motivational factors. This is in contrast with the idea of a general impairment in a transversal cognitive capacity such as working memory. I propose that the persistence of the delusional belief is the result of a normal process that has at its core the avoidance of dissonant or negative conceptions about the self. I suggest, more specifically, that in this fragmented and psychofunctional system, the patient suffering from delusions might resist counterevidence when it causes distress or goes against core beliefs about the self.

First, the delusional belief is formed as a prepotent response to an abnormal experience (as in Davies & Egan, 2013). The belief is then compartmentalized in a way that inconsistent beliefs in other fragments remain unaffected. This explains the lack of integration of the delusion and the behavioral profile of the subject, who only sometimes acts in accordance to the delusional belief. Again, the central issue is the persistence of the delusional belief when confronted with contradictory evidence. That is, when faced with background beliefs of the low probability of a look-alike stranger replacing one's relative, and new evidence obtained in interacting with the alleged impostor (such as that he has the memories of the relative, seem to abide by its habits, and so on). Following Bendaña and Mandelbaum's belief system, the psychological immune system might be at play. First, among the core beliefs about the self, there is also the one of being a mentally healthy person (namely, *being sane*). When confronted with internal or external evidence that goes against the content of the delusion, the patient might go through the following chain of reasoning:

> P1) I have the *feeling* that this man, who looks exactly like my husband, is not my husband.
>
> P2) Only someone crazy would have the feeling that he is not in front of his husband when being in front of him.
>
> P3) I am not crazy.

C) This man is not my husband.

This line of reasoning would make him persist in the delusional belief in the face of counterevidence. In this example, which parallels the Marine's case of the previous section, the core belief about being mentally healthy would impede a deeper consideration of the evidence against the delusional belief. The desire to avoid simultaneously believing that one is and is not crazy mediates this process. No matter how strong the evidence counter to the delusional belief (from doctors and friends' opinions), the belief about being sane is stronger and will not be overridden. Crucially, this line of reasoning would only result in this conclusion when P1 is the case. That is, when the subject is having, or has recently had the experience of seeing a relative and feeling that he *looks like* her relative, but it does not *feel like* him. If this feeling were not present, the line of reasoning would be void.

b) Psychological resistance to fragment coactivation and merge

Given that both the delusional belief and its opposite seem to operate over time, the proposal also needs to explain the persistency over time of contradictory beliefs residing in coactivated fragments—as we have reason to believe happens in Capgras patients. This is especially the case because, if coactivated, these fragments should be made consistent. For instance,the belief "if a person is physically identical to my husband, he is my husband" might be part of the patient's background beliefs. This belief is incompatible with the impostor hypothesis and, in fact, might guide some patient behaviors, such as accepting living with the impostor. Even though Bendaña and Mandelbaum's model allows for inconsistency among beliefs stored in different fragments, it also claims that when two fragments are coactivated they would merge and, therefore, be rendered consistent. It is expected that if the subject activates the fragment containing the delusional belief "the man identical to my husband is not my husband" and another fragment containing the inconsistent belief "the man identical to my husband is my husband", one of these beliefs should be eradicated. The authors posit a principle that, in these cases, would decrease the likelihood that inconsistent beliefs are coactivated:

> "Fragmentation allows for the sequestering of inconsistency, but how does the mind actually reduce the likelihood of coactivating inconsistent beliefs? We hypothesize that the mind accomplishes this by operating in accordance with the 'let sleeping dogs lie' principle (McDermott 1987). Roughly, the principle is one of cognitive economy: one conserves cognitive energy unless spurred on by an external event or command. Applied to Fragmentation, the principle dictates thata fragment remains quiescent unless a) a search is triggered for its specific heading, and b) once that heading is located, searches cease. As long as inconsistent beliefs are housed in separate fragments, a sleeping-dogs principle dramatically decreases the likelihood of coactivating the inconsistent beliefs" (Bendana and Mandelbaum 2021).

However, in such pathological cases, there are external pressures to coactivate fragments containing the delusional belief and its opposite, for example, under the insistent prompts of doctors. As a consequence, the patient should lose one of the two inconsistent beliefs. Nevertheless, this does not occur because both beliefs guide the patient's behavior in a context dependent way over time.

Bendaña and Mandelbaum's fragmented model could accommodate this phenomenon by positing a psychological principle of resistance to fragment coactivation. Coactivating fragments and having to render them consistent can cause psychological distress and threaten core beliefs about ourselves. The coactivation of a fragment containing the proposition "there are oranges in the fridge" and another with "there are no oranges in the fridge" would not cause any harm. The fragments would merge and one of the beliefs would be eradicated. However, the activation of the delusional belief "this man is not my husband" and the opposite "this man is my husband" would cause distress since it threatens the belief in one's sanity. In this case, the patient would realize that holding both beliefs at the same time—as he had been doing for some time—is pathological, which would undermine the core belief of being a sane person. It is important to keep in mind that the subject has reasons not to doubt his mental health: Cognitive functioning is maintained in these patients, and the delusion concerns only a single topic. Because of this, there might be a subpersonal prediction of what the coactivation of these two fragments would amount to. If psychological distress is anticipated, coactivation of these fragments would be resisted. This principle agrees with the fact that patients suffering from delusions frequently attempt to change the topic of conversation and exhibit discomfort when confronted with inconsistencies between their delusional beliefs and other background beliefs (Halligan and Marshall 1995). This interpretation is also in line with some descriptions of the phenomenon: "It seems as if thenew information does not even enter the deluded subject's belief system as data that needto be explained" (Coltheart et al. 2010: 280). Motivational factors would impede fragment coactivation when externally invited to do so if this coactivation generated distress and affected the core beliefs of the subject.

The Capgras delusion can be explained in this fragmented psychofunctional framework without positing additional cognitive impairments, which is consistent with the current evidence. This goes in line with accounts that only one deficit—the one concerning the autonomic system—is needed to account for the delusion (Maher's 1974). Furthermore, it agrees with the claim these accounts make that mechanisms of belief fixation operate normally in patients with these delusions. Modelling the Capgras delusion and other monothematic delusions in a fragmented psychofunctional system of belief also aligns with many of the characterizations of delusions made in the recent years. For example, Bortolotti's (2010) suggestion concerning the fact that the epistemic features shown in delusions are characteristic of many of our everyday beliefs and are not limited to pathologies of the mind. It is also consistent with the connection pointed out in the last years between delusion and self-deception (Bayne and Fernández 2009). In conclusion, the questions concerning the persistence of the delusional belief can be better answered by a fragmented belief system that considers motivational factors. At the same time, this framework can explain the circumscribed character of the delusion and its behavioral profile.

3.5. Conclusions

My aim in this chapter was to point to inconsistencies in other accounts of delusion and provide an alternative. First, I provided criticisms of Currie's and co-authors metacognitive account of delusion based on new grounds. Namely, if delusions were the kind of imaginings they propose, they should be responsive to evidence, which is not the case. Then, I have provided an alternative account using a fragmented system of belief. I first considered two fragmented systems of belief. Davies and Egan's account (2013) is unsatisfactory since it either leaves the persistency of the delusional belief when confronted with counterevidence unexplained or explains this persistency at the expense of leaving the patient's inconsistent behavioral profile unexplained. Bendaña and Mandelbaum's (2021) fragmented and psychofunctional system of belief offers better explanatory resources in modelling the Capgras delusion. It accounts for the adoption of the delusional belief and its persistence by appealing to motivational principles governing belief updating. This fragmented model can also explain context sensitivity. It also has the resources to account for the circumscribed character of the delusion and the patient's context-sensitive behavioral profile. Furthermore, by positing psychological motivations in belief updating and revision, it does not need to postulate cognitive impairments in the patient for which goes against current evidence.

Chapter 4

There is no smoke without fire: The Simulation Theory of memory and the phenomenology of remembering

Abstract. The Simulation Theory of memory states that to remember an episode is to simulate it in the imagination (Michaelian, 2016a, 2016b), being memory thus reducible to the act of imagining. This chapter examines Simulation Theory's resources to account for our ability to distinguish episodic memory from free imagination. The Theory suggests that we can reliably do so because of the distinctive phenomenology episodic memory comes with (i.e., a *feeling of remembering*), which other episodic imaginings lack. In this chapter I raise two objections to how the feeling of remembering is engineered in the theory, followed by an exhaustive exploration of the theory's resources to ground the mechanism underlying the raising of such feeling. I conclude that the Simulation Theory cannot simultaneously defend the simulational character of episodic memory and ground our ability to discriminate between memories and imaginings.

4.1. Introduction

Our mental life relies heavily on visual experiences. Its relevance goes beyond the here and now that characterizes perception. Not only can we *see* a trapeze artist somersaulting at the circus, but we can also visualize it "in the mind's eye" when *remembering* it. And even if we have never seen a trapeze artist, we can *imagine* it. Due to its episodic and almost sensory form, these phenomena have been named quasi-perceptual memory and imagination (Macpherson and Dorsch, 2018)[35]. Traditionally, this kind of memory and imagination has been considered two distinct kinds of mental activity (Bernecker, 2008). Recent theories, however, claim that the difference between them is of degree, not of kind (Michaelian, 2016a, 2016b; De Brigard, 2014a; Hopkins, 2018). According to Michaelian's Simulation Theory (2016a, 2016b; Simulationism or STM henceforth), to remember an episode is to simulate it in imagination. This raises the following relevant question: when an individual is entertaining certain imagistic content, how can she tell whether she is remembering or engaging in free imagination? This chapter will address the ability of STM to give a proper answer to this crucial question.

The traditional answer to this question in the literature has been to appeal to the distinctive phenomenology of memory. Episodic memory and experiential imagination are said to *feel* different, even though both imply the visualization of a scene *in the mind's eye*. STM also appeals to this phenomenological character of episodic memory. Michaelian claims that episodic memories come with a distinctive phenomenology, a *feeling of remembering*, that accompanies the episodic representation and allows the subject to identify memories as such (2016a, p. 235). Nonetheless, appealing

[35] As the example used shows, the memory and imagination that concern us here are episodic and experiential, rather than propositional. An episodic memory (e.g., the memory of *swimming* at the River Ouse on a summer morning in 1994) can be contrasted with a propositional or semantic memory, which does not include imagery and consists of the retention of a particular belief (e.g., the belief that the River Ouse crosses the county of East Sussex). Something similar happens in the case of imagination, which has an experiential and propositional variant (e.g., imagining *submerging* in the River Ouse vs. imagining a state of affairs being actual, such as that Caesar's troops crossed the River Ouse during the Gallic War -something that, in fact, never happened). Both episodic memory and imagination concern imagery in all modalities (vision, hearing, taste); for explanatory purposes, examples here will focus on the visual.

to this *feeling* from a simulationist stance demands a more thorough explanation. If the process that brings memories and *other* imaginings to mind has the same features, how does memory have a distinctive phenomenology? I will first evaluate how STM addresses this issue. Then, I will explore the resources of Simulationism to account for the distinctive phenomenology of memory without undermining its central ontological claim: to remember is to imagine. In the same way the presence of smoke implies the combustion process that gives rise to it, the emergence of a differential phenomenology for memory requires an underlying differential mechanism. The question to be answered is the following: Can Simulationism ground a phenomenological difference between episodic memories and (*other)* imaginings?

Section 2 reviews the central claims of Simulationism. Section 3 focuses on the phenomenology of memory as stated in STM. In section 4, I raise two objections to the way memory's phenomenology is described to arise according to the theory. To amend the shortcomings raised, section 5 explores possible underlying causes of the feeling of remembering in a simulational paradigm. As a preview of the results, I find that none of them constitutes a solution to the problems raised. I will then conclude that, unless amendments are made, Simulationism cannot simultaneously defend its central claims and ground the ability to distinguish episodic memory from "other episodic imaginings".

4.2. The Simulation Theory of memory

As previously said, Simulationism claims that episodic memory is the result of our imaginative capacities put to the purpose of constructing—namely, simulating—a representation of an episode of the personal past. In Michaelian's words (2016a, p. 60)

> "Fundamentally, on this view, remembering is generative, not preservative: it is not a matter of preserving a representation but

> rather of constructing, on the basis of stored information originating in a variety of different sources, as well as information available in the subject's current environment, a *new* representation of a past episode. In short, remembering is a matter of imagining or simulating the past."

The theory is empirically motivated by two well-established discoveries. First, in attempting to remember an episode, we often combine it with information obtained from other episodes (e.g., Brainerd and Reyna 2005, Loftus 2005) and from other sources (e.g., testimonial information[36]; Meade and Roedinger, 2002). Because of this, episodic memory is seen as more constructive than was initially posited by preservationist models, which described it as a process of storage and retrieval (Dummet, 1994; Audi, 1995). The second discovery is that remembering and imagining share a common neurocognitive structure (Addis et al., 2007; Szpunar et al., 2007, Mullally et al., 2014). Both phenomena motivate the postulation of the *Episodic Construction System,* devoted not only to the simulation of past episodes but also to a wide range of imagined episodes. Among these are episodic future thought (Szpunar, 2010) and episodic counterfactual thought (episodic imaginings about what could have happened in the past; De Brigard, 2014a). Episodic memory, therefore, is one example among many episodic imaginings (Michaliean 2016a, p. 111). ST states that a subject S remembers an episode *e* if and only if (2016a, p. 107):

1) *S* now has a representation R of *e*
2) R is produced by a properly functioning episodic construction system which aims to produce a representation of an episode belonging to *S*'s personal past.

[36] Namely, information about our past received through communication with other agents.

In emphasizing the constructive character of memory, the theory dispenses two classical requirements in the philosophy of memory. For the simulationist, an episodic memory can be entirely constituted by information originated in similar episodes or coming from testimonial sources, if it represents an event of the personal past, and it is produced by a properly functioning episodic construction system. In this regard, STM discards the Causal Condition for memory (Martin and Deutscher, 1966), in which remembering requires a continuous causal connection from the subject's original experience of the event to her retrieved representation of such event.[37] In the following case, according to STM, Felicia would be remembering:

> **[1] Forgetful Felicia and an afternoon at the circus:** At the age of six, Felicia goes to the circus with her parents and brother. Due to an accident, she loses all her memories of this event. Years later she is told about the episode by her brother. Later, she forgets having been told and based on the testimonial information given by her brother, she imaginatively represents the event in her mind: the trapeze artist dancing on an elephant, the smell of popcorn mixed with the smell of animals; her excitement at all.

As in **[1],** STM claims that we can remember experienced episodes fully based on non-experiential information: episodic memories do not need to draw on information originating in the experience of the remembered episode *at all*.[38] Moreover, a second classical requirement for memory that STM discards is the Previous Experience Condition. This condition states

[37] STM dispenses with this condition for the following reason. Knowing as we do that remembering involves the reconstruction and incorporation of information from many sources beyond experience, it is unjustified to stipulate that a minimum percentage of this information must come from the original experience of the episode via an appropriate causal link. Additional requirements in the literature concerning this condition, are the fact that this causal chain is appropriate only if it goes via a memory trace. For simplicity, here I will focus on the most important conditions denied by Simulationism.

[38] By non-experiential is meant information that we have not acquired first-hand, such as testimonial information.

that for a subject S to remember an episode *e*, S needs to have experienced *e*. On the contrary, according to Simulationism, we can remember episodes that we have not experienced[39], as it happens in the following case:

> **[2] Felicia at the age of two:** Suppose a case identical to [1], with the only exception that, in this case, Felicia went to the circus at the age of two, too young to count as having *experienced* [40] the episode. She is later told about this episode by her brother, forgets about having been told, and lately, on the mere basis of the testimonial information given by her brother, she imaginatively represents the event in her mind.[41]

Because for the simulationist no percentage of the content of an episodic memory has to be retrieved from the original experience—as it happens in [1] and [2]—, Michaelian claims that in these cases episodic memory generates new beliefs along with its content (as it happens in perception). In his terms, episodic memory is a *radically generative* epistemic source: it can not only justify beliefs but also be the very source of this justification, by providing the contents that justify the formation of these beliefs.

In what follows, three central claims of STM will be of use throughout this work: 1) that to remember is to imagine our personal past, 2) that memory is produced by the episodic construction system, and 3) that episodic memory is a radically generative epistemic source. Let us now focus on Michaelian's answer to our initial question regarding people's ability to distinguish memory and imagination.

[39] To remember, in STM, it is sufficient that the episode we represent belongs to our personal past: we need not have experienced it. This surprising claim is explicitly stated by Michaelian: "the recreative character of remembering requires us to abandon the idea that things remembered must be things formerly perceived or known" (2016a, p. 60).

[40] Michaelian adopts a narrow notion of experience in [2], but these need not concern us here.

[41] The counterinitiative fact that cases as [1] and [2] are counted as instances of episodic memory has recently been discussed in the literature (McCarroll, 2020) and will not be the subject of debate here. Examples are given to characterize the theory and will be referred in the chapter.

4.3. Tracing back phenomenology in Simulationism

Despite emphasizing the non-reproductive character of episodic memory, Michaelian defends its overall reliability, namely, its epistemic status as a process through which subjects form accurate beliefs about the past. The main reason why he sees the constructive character of episodic memory as no threat to the reliability of remembering is the following: He claims that when the episodic construction system generates an episodic memory, this memory comes with a specific phenomenology. This phenomenology is exclusive of episodic memory, and the other range of episodic constructions (e.g., daydreaming, episodic counterfactual thought, or episodic anticipation) lack it. Thanks to the phenomenology of remembering, we can reliably distinguish remembering from *other* imaginings [42]. According to Simulationism the *feeling of remembering* is the crucial element in ensuring that episodic memory is reliable, despite the facts that it is constructive, frequently based on non-experiential information, and sometimes concerning non-experienced episodes. The content of this *feeling* emerges in consciousness as "*this representation is a representation of an event from my past*" (Michaelian 2016a, p. 235)*,* allowing the subject to discriminate—most of the time, reliably—memory from *other* episodic imaginings.

Appealing to the subjective character of mnemonic contents—to the way they *feel*—as the marker that allows subjects to identify memories first-personally is a common occurrence in the literature. This qualitative feature or *what-it-is-likeness* singular of episodic memory has received many characterizations. James (1890, p. 650) refers to memories as having a "*feeling of warmth and intimacy*"; Russell's *pastness* (1921, p. 163) claims that the contents of episodic memories are accompanied by a *feeling of familiarity*; and Tulving (2002, p. 6) emphasizes that when we remember, we seem to *re-*

[42] This claim does not preclude that on some occasions, subjects will erroneously judge that they are remembering when they are imagining (and vice versa). Nonetheless, the common practice reflects that in most cases, we correctly determine whether we are remembering or imagining -at least in the case of healthy subjects. Therefore, we tend to trust this capacity, and turn to it when we want to know our past.

experience past episodes. Other characterizations include Dokic's "*feeling of knowing*" (Dokic, 2014), Fernàndez's "*feeling of ownership*" (2019), and Martin and Hoerl's "*feeling of particularity*" (Martin, 2001; Hoerl, 2001).[43]

However, if episodic memory and imagination result from the same constructive and additive process and the system producing them is the same, why would the process elicit the *feeling of remembering* only when the represented episodes are part of our personal past? What is the mechanism underlying the qualitative distinctiveness of episodic memories? If Simulationism wants to ensure the reliability of memory despite its imaginative nature, it must give a detailed account of how the *feeling of remembering* originates. In its formulation, the feeling originates as follows (2016a: 232):

> "Given its simulational character, remembering would be unreliable and therefore maladaptive absent the subjective dimension, for agents would be unable to reliably distinguish among different forms of episodic imagination. If an episodic constructive process is classified as self-oriented, past-oriented, and actual rather than counterfactual, it is judged to be an instance of remembering—the agent has a *feeling of remembering*"

In short, in STM the phenomenology of memory emerges if the episodic construction is self-oriented (i.e., autonoesis), past-oriented (i.e., chronesthesia), and taken to be actual (i.e., actuality). If these three conditions are met, the episodic representation brings, in conjunction with the contents represented, the feeling of remembering (see Figure 1). In virtue

[43] The debate concerning the best characterization of the phenomenal marker that allows the subject to distinguish remembering from imagining is still alive (Byrne, 2010; De Brigard, 2017; Teroni, 2017), although we need not engage in this debate for our purposes in the chapter. On the other hand, skeptic views about episodic memory having a distinctive phenomenology are rare (see Hopkins 2018 and Hoerl 2019). I take the phenomenology of memory as intuitively plausible and empirically well-established, and for reasons of space, I will not question it in the chaper.

of this phenomenology, the subject judges that she remembers and, therefore, that the episode represented took place in her past.

Concerning the characterization of these conditions, some clarifications are necessary. The first condition, *autonoesis* [44] refers to the fact that the episode needs to be self-oriented. One can experientially imagine oneself seeing the Pyramid of Khafre, but one can also imagine being Howard Carter seeing that pyramid: only in the first case is the episode self-oriented. The second condition, *chronestesia* concerns the temporal orientation of the episode, which needs to be past-oriented as opposed to present- or future-oriented. One can imagine oneself at the age of six exploring nature with a brother, but one can also try to anticipate the future and imagine oneself at the age of 60 exploring nature with a grandson; only in the first case, the episode is past-oriented. Concerning the last condition, *actuality,* the representation needs to be taken as actual (as having occurred) instead of counterfactual (something that could have occurred). An example of an *actual* event is one's memory of yesterday morning at the beach; an example of a counterfactual simulation is imagining what would have happened if one had gone to the park instead.

Due to the allusion to necessary precursors to the feeling of remembering, I will label this explanation the "Phenomenological Precursors Strategy". In the next section, I will give two arguments to show that this strategy fails to ground the feeling of remembering.

[44] Michaelian uses autonoesis to describe the episode as self-oriented (represented from the perspective of the subject). Other, more compromised, uses of the term can be found in the literature (e.g., Tulving 1985)

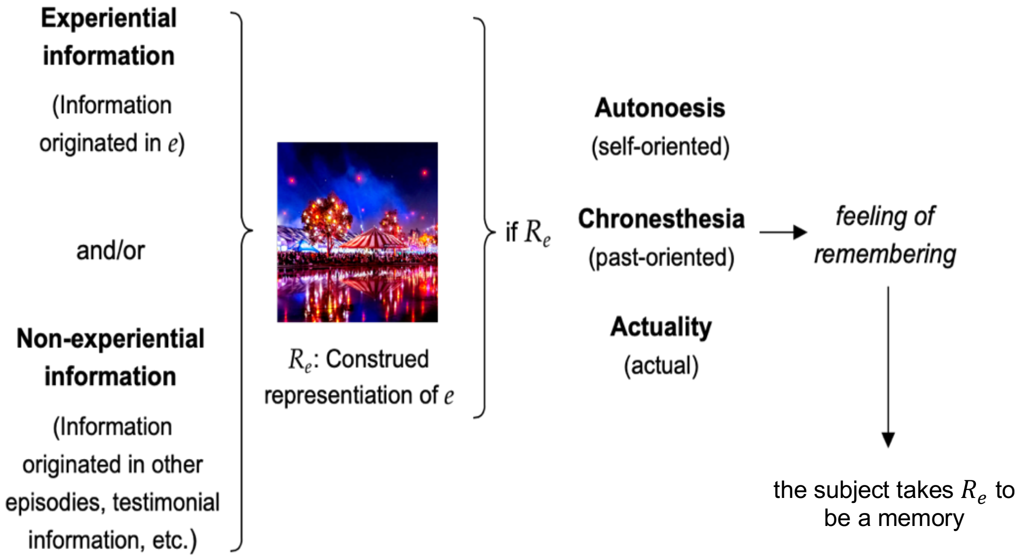
Experiential information
(Information originated in e)
and/or
Non-experiential information
(Information originated in other episodies, testimonial information, etc.)
R_e: Construed representiation of e
if R_e
Autonoesis
(self-oriented)
Chronesthesia
(past-oriented)
Actuality
(actual)
feeling of remembering
the subject takes R_e to be a memory

Figure 1. The genesis of the *feeling of remembering,* as stated in Simulationism (*e* stands for the event of the personal past that is being remembered).

4.4. Two objections against the Phenomenological Precursors Strategy

We have just seen how STM gives rise to the feeling of remembering. I will now argue that the decomposition of the feeling of remembering into its alleged three precursors does not amount to an explanation of the mechanism that produces it. I will show this using two different arguments. The first concerns the fact that the Phenomenological Precursors Strategy begs the main question we are trying to answer here (section 4.1). The second suggests that the three precursors are insufficient for the feeling of remembering to emerge (section 4.2).

4.4.1. Actuality: an unexplained explainer

Michaelian acknowledges that episodic memory and other forms of episodic imagination partially share their phenomenology. This fact sits well with the predictions of the Simulation Theory. Given that memory and imagination are produced by the same episodic construction system, phenomenological similarity between them is to be expected. For instance, the self-oriented condition (i.e., autonoesis) and the past-oriented condition (i.e., chronesthesia) can be present in other episodic imaginings apart from episodic remembering. The following episodic counterfactual imagining provides an example:

> **[3] Felicia and her counterfactual past:** One Sunday afternoon, Felicia entertains herself imagining how the visit to the circus when she was six would have been if, instead of going with her parents, she had gone with her more permissive grandparents. Probably, they would have let her climb on the elephant. She imagines things from up there: the closeness with the head of the adjacent giraffe, the touch of the elephant's back; and so on.

In [3], Felicia constructs an episode that is self-oriented (autonoesis) and past-oriented (chronesthesia). The fact that she takes the representation to be counterfactual rather than actual, impedes the raising of the feeling of remembering and prevents her from considering the task to be remembering. This, in turn, prevents her from forming beliefs about the contents represented (such as that she climbed on an elephant at the age of six). This example also shows that voluntarism can be claimed about *autonoesis* and *chronesthesia*. That is, it is always possible to ascribe at will the *self-oriented* and *past-oriented* conditions to an imagined episode, converting it into an instance of episodic counterfactual thought. However, and crucially, according to STM the feeling of remembering cannot be induced at will, since this would mislead the subject about the mnemonic nature of the representation. The *actuality* condition secures this: one cannot take the contents represented as *actual* at will. *Actuality* distinguishes episodic memory from episodic counterfactual thought: both constructions are self-oriented and past-oriented, but only the contents of the former *e* are classified as *actual*. This shows that the modal condition (actuality), absent in [3], is the crucial condition in accounting for the rise of the feeling of remembering.

Despite this, the simulation theory remains silent about how episodic representations are classified as *actual*.

When and how are the episodic contents taken as *actual*? What is it that invites the subject to take them as such? STM posits the *actuality* condition as part of the explanation for the emergence of the feeling of remembering, but the origin of this crucial condition is not described in the theory, remaining an *unexplained explainer*. The decomposition of the feeling of remembering into its alleged precursors—autonoesis, chromesthesia, and actuality—leaves an explanatory gap, since the origin of the crucial precursor, *actuality*, remains ungrounded and mysterious, leaving our main question unanswered. Although not considered by Simulationism, in section 5, I shall explore potential candidates for grounding the phenomenology of memory in STM. Before that, we shall consider the second argument against the Phenomenological Precursors Strategy.

4.4.2. On the insufficiency of *autonoesis, chronesthesia, and actuality*

As seen in [3], we can imagine episodic counterfactual episodes at will. Since these counterfactual episodes are self-oriented (autonoesis) and past-oriented (chronesthesia), it follows that the conditions of autonoesis and chronestesia can be met at will. For example, one can imagine what Juana de Arco saw and felt when leading the siege of New Orleans at seventeen (picturing the battle in front, hearing the sound of the horses, and feeling the fear of imminent death). However, one can also orient the episode to oneself (*autonoesis*) and imagine this counterfactual past: one commanding a siege at seventeen. Luckily, one would not be able to take the contents represented as "actual", and therefore, the feeling of remembering would not emerge when imagining one's belligerent counterfactual past. However, should we not be able to bring up the feeling of remembering at will in some cases? Let us consider the following case:

> **[4] Felicia, the academic.** Having learned that Simulation Theory predicts that if an episodic imagining is self-oriented, past-oriented, and taken as actual, the *feeling of remembering* will rise, Felicia decides to put this to the test. She asks her older brother—whom she takes to be a reliable source about her childhood—to give her detailed information about an event they experienced together when she was two years old, and about which she remembers nothing. After compiling the information, she imagines the episode orienting it at herself (autonoesis) and to the past (chronestesia). Furthermore, she takes it as actual—instead of counterfactual—since she has construed it based on reliable information and believes it to have happened.

In cases like [4], I claim that the three conditions are met (i.e., autonoesis, chronesthesia, and actuality), but no *feeling of remembering* accompanies the episodic representation. The feelings surrounding this kind of constructed episodes are closer to the ones of episodic counterfactual thought. Given that, [4] posits a counterexample to the sufficiency of the three precursors conditions for the raising of the feeling.

Simulationists can reply that phenomenological intuitions are slippery, but at least when firstly imagined[45], experiential imaginings like [4] are accompanied by the same phenomenology of strangeness and remoteness as those of counterfactual imagination in [3].[46] If in [4] Felicia

[45] It may happen that, after repeatedly imagining an episode from our childhood that has been narrated to us by testimonial sources, we end up generating that phenomenology of memory at the umpteenth attempt. If this is the case, it would be an instance of phenomenon of imagination inflation (Garry et al., 1996), in which we mistake imagination for memory. But this is not the case in [4].

[46] Mahr (2020, p. 8) shares this phenomenological intuition: "You might have had too much to drink one night and therefore wake up the next day without remembering anything of what occurred. Your friend, who was with you at the time, however, tells you in a lot of detail what occurred, namely, that you got into an argument with the barman about how many drinks you had. Now, you might be able to simulate fairly accurately this specific, past event, which you will take to have actually occurred and which you will take to have occurred to you personally. You will, however, still not take yourself to remember the event."

takes the contents to represent an event of her personal past, this is just because she takes her brother as a reliable source, not in virtue of any *feeling of remembering*. This is not an isolated case but rather the norm for cases in which testimonial information is consciously incorporated. Episodes we construe based entirely on the conscious incorporation of new testimonial information (even if we construe the episode in imagination as past-oriented and self-oriented) seem to have a phenomenology more similar to counterfactual imaginings and do not seem to come accompanied by the feeling of remembering.

The STM advocate might object that, since [4] is entirely based on the *conscious* incorporation of testimonial information, this is not an instance of episodic remembering. When discussing the incorporation of reliable testimonial information in episodic memories, Michaelian uses only examples in which such incorporation is *unconscious* (namely, the person remembering is not aware of it). For this reason, it could be replied that because in [4] the subject consciously forms a representation that incorporates testimonial information, the representation in question is not the product of the episodic construction system. It could also be denied that in such cases the episodic construction system is functioning properly (Michaelian, personal communication). If this was the case, instances like [4] would not be remembering because the metaphysical conditions of the theory would not be met.

One may reply to this line of thought in two ways. First, even if [4] was not an instance of episodic memory, it would still be a case against the sufficiency of the three precursors for the emergence of the feeling of remembering. That is, a counterexample to the sufficiency of the three conditions (autonoesis, chronestesia, and actuality) for the feeling to rise. Second and more importantly, from a simulationist stance, it is unmotivated to deny that in cases like [4] the memory system is involved and functioning correctly. If memory is part of the episodic construction system, such system 1) frequently receives inputs that are introduced consciously and 2) frequently construes episodes based exclusively on the conscious incorporation of information. This is the case, for example, of episodic anticipation, in which many times we consciously set up an imagined scenario. Why would the simulationist posit an asymmetry in the functioning of the episodic construction system, accepting as input conscious information in the case of episodic anticipation but not in the case of episodic memory? Positing this input asymmetry seems to go, in fact, against the existence of the system, as the kinds of inputs that a cognitive system is sensitive to are one of the main criteria for individuating it. On the other hand, it is unreasonable to claim that the conscious incorporation if information is less optimal than its unconscious incorporation. The main difference between the two is that, in the case of conscious incorporation, the subject is often able to check the source of information, which can only increase the reliability of the process.

Because in Simulationism the phenomenology of memory is not a necessary condition for remembering, the previous objection does not affect the metaphysical formulation of the theory. Concerns about how the subjective dimension of episodic memory is engineered in the theory are relevant if it wants to claim that memory, although constructive, is reliable because of its phenomenology. As presumably shown, the theory's predictions concerning the emergence of this *feeling* do not align well with phenomenological facts and fail to do so systematically. The mechanism proposed by STM to underlie the phenomenology of memory has been proven ungrounded (4.1) and its conditions insufficient (4.2). At this point, Michaelian could adopt a skeptical strategy and claim that the phenomenological dimension is not indispensable for the reliability of memory. Let us now see why this is not a suitable solution to the problem.

4.4.3. On the indispensability of phenomenology for reliability

Faced with the previous objections, STM could leave out the phenomenological dimension of episodic memory. It could be claimed that the theory is committed to memory being reliable, not to memory belief-formation being reliable. That is, it could defend that these kinds of reliability are independent, and that reliably remembering does not imply reliable memory belief-formation. For example, a subject with highly inaccurate metacognition might reject a lot of accurate memories (Michaelian, in conversation). Therefore, before exploring some candidates for grounding the phenomenology of memory in Simulationism, I will motivate the indispensability of phenomenology for the reliability of memory.[47]

It seems to be a desideratum of any theory of memory that defends the overall reliability of memory to also account for reliable memory belief-formation. It is in virtue of identifying the imagistic contents as episodic memories ("I am remembering") that subjects form beliefs about the contents represented ("This happened"). For example, when entertaining the images of grandma disguised as a dinosaur at a Carnival party, if I take this construction to be a memory, I will form the belief "Grandma came to that Carnival party" when visualizing such scene in the mind's eye. If, by the contrary, I take it to be a product of free imagination, I would not form this belief. The recognition of memories as such is an indispensable last step for them to play the functional role they play, and the system producing memories should explain part of this recognition process. The following analogy will be helpful in understanding why simulation without appropriate phenomenology would lead to unreliability. Imagine a blind master perfumer whose purpose is to make a perfume of lilies. When she goes to the garden,

[47] Here, the focus has and will be on the case in which phenomenology is the central feature for distinguishing both faculties. Nonetheless, what has been said applies to other non-phenomenological markers posited as the mechanism by which we first-personally distinguish memories and imaginings. That is, if a theory denies that memory comes with a particular phenomenology, it will still need to ground the first-personal memory judgement about a certain episodic representation (namely, the fact that we tend to recognize memories as such).

her hands duly select the lilies, distinguishing them from the roses—most of the time reliably. After carrying out all the proper steps to obtain the lily perfume, the master perfumer smells it. What a great mistake it would be if the lily perfume smelled sometimes of lilies and sometimes of roses! The whole process of distilling the perfume, no matter how careful, would be useless if it did not end up evoking in the perfumer the phenomenology of the smell of lilies that allows her to identify it as such, and consider it finished.

If we could not identify memories as such and distinguish them from free imagination and we were constantly confusing one for the other, memory would be of little use. This faculty would continually mislead us as to what happened in our past and would not be a faculty to trust. Luckily, this is not the case: experience shows that episodic memory is a reliable process most of the time—at least in healthy subjects—, and we continuously turn to it when we want to obtain information about the past. So, explaining how we distinguish episodic memory from other forms of episodic imagination when entertaining certain types of episodic content is ineludible for the simulationist if he wants to claim that episodic memory is reliable.

In the next section, several candidate mechanisms for grounding the phenomenology of memory will be considered along with their compatibility with the central claims of Simulationism. Unfortunately, the conclusion will be that none of them can be taken by the simulationist without renouncing some central claims of the theory.

4.5. Grounding memory's phenomenology in Simulationism

In the following sections, I shall explore several candidates that the simulationist could appeal to in grounding the phenomenology of episodic memory. These candidates have been the predominant ones in the literature on memory first-person markers (Byrne, 2010; Teroni, 2017; Perrin et al., 2020). They can be divided into three groups: procedural, doxastic, and intentional. In each case, the postulated mechanism is incompatible with some central claim of Simulationism. The mechanism underlying the distinctive phenomenology of memory remains unfilled in STM, and it does not seem easy to find an STM-compatible candidate. Without such a mechanism, the reliability of memory defended in STM remains ungrounded, deeply undermining the theory's explanatory power.

4.5.1. Procedural features

Those who take memory and imagination as different mental processes—*discontinuists*—can easily account for the distinctive phenomenology of memory and imagination: different processes can have different phenomenological outputs. Once an ontological difference between both processes is assumed, the distinctive phenomenological output of memories can be explained by appealing to the nature of the process that gives rise to them. For instance, some causal accounts of memory (Bernecker, 2010) endorse the existence of memory traces that encode and preserve information about an event over time (De Brigard, 2014b; Robins, 2017; Werning, 2020). These memory traces are said to be causally operative in producing a memory representation. This distinct feature of memory—the activation of such traces—, therefore, could lead to the overall phenomenology:

a) ***<u>Nature of the process or the subpersonal detection of its features:</u>*** What causes the feeling of remembering is a differential feature—or a subpersonal detection—of the process giving rise to memories.

But in Simulationism the process that gives rise to episodic memories and other episodic imaginings is the same, and the existence of information originated in the remembered event—or any memory trace of it—is not necessary (as seen previously in cases [1] and [2]). This makes it implausible to ground the *feeling of remembering* in a distinctive feature of the mnemonic process, nor in its subpersonal detection.

However, one counterargument by STM could be that, in grounding the phenomenology of memory, there is no need for the process that give rise to memory and other imaginings to be different in nature, but rather different in their average features [48]. It could be claimed that although the process that simulates episodic memories and *other* experiential imaginings is the same (in both cases constructive and additive), there are average differences in the running of that process. These differences, in turn, are subpersonally detected, and this detection contributes to the overall phenomenology of memory. Authors like Dokic (2014) have proposed that the subpersonal detection and interpretation of average cues such as fluency might have, as a result, the characteristic feelings that accompany memories. The phenomenology of memory could be, in the case that concerns us, the result of the subpersonal sensibility at the personal level to the average features of episodic memory[49]. Therefore, it could be claimed that although the process underlying memories and other imaginings is the same, the procedural fluency of memories is on average higher than that of other imaginings. The subpersonal detection of this feature, in turn, would give rise to the *feeling of remembering.* Other features such as the ease of generation could also be appealed to (Wittlesea & Leboe, 2000). Hence, a second candidate STM could allegedly endorse is the following:

- **b)** ***Subpersonal detection of average features:*** What causes the feeling of remembering is the subpersonal detection of average features of the process giving rise to memories (e.g., the higher fluency or ease of generation in episodic memory compared to other imaginings).

[48] Michaelian suggests this at some point (2016a, p. 196).

[49] See Whittlesea, 1997, p.219 and Koriat, 2007, p. 298 for similar claims.

However, it seems that STM could not coherently endorse such a mechanism for several reasons. STM's characterizes memory as a highly reconstructive and additive process, inconsistent with invoking heuristics relying on average differences such as fluency. If the feeling of remembering was the result of the subpersonal detection of the process as fluid, then highly construed memories—which are taken to be frequent in the simulationist framework—would not be subpersonally detected. Therefore, numerous instances of episodic memory would lack the phenomenology of memory. The absence of the feeling of remembering, in turn, would lead the subject to misjudge these memories as counterfactual imaginings or to suspend judgment about them. Since these memories are very frequent, the overall reliability of episodic memory would then be under threat. Such an argument runs as follows (P1 is the candidate under consideration. P2 is a central claim from Simulationism and one of its central motivations for its ontological thesis. P3 concerns the nature of procedural fluency, and P4 follows from P2 and P3):

> **P1**) The *feeling of remembering* emerges only when high procedural fluency is subpersonally detected.
>
> **P2**) Many episodic memories are highly constructive and additive in nature.
>
> **P3**) The more constructive and additive an episodic construction is, the less the procedural fluency of the process running it.
>
> **P4**) Many episodic memories have low levels of procedural fluency.
>
> **C**) Therefore, many episodic memories lack the *feeling of remembering* (From P1 and P4).

As the argument shows, emphasizing the contrived and additive character of memory to the point of equating it with imagination is inconsistent with simultaneously emphasizing memory's fluency and ease of generation as a phenomenological marker. The procedural strategy, both concerning the nature of the process and its average features, is not compatible with some central claims of STM. More specifically, *(a)* clashes with the ontological claim that subsumes memory within imagination. On the other hand, *(b)* along with STM claims on the nature of episodic memory, leads to the conclusion that many episodic memories lack the *feeling of remembering,* and therefore would not be recognized as such by the subject having them (which undermines the reliability of memory).

4.5.2. Doxastic coherence

An alternative option from the simulationist standpoint would be the following: the feeling of remembering could emerge after a process of comparison between the episodically represented contents and propositional beliefs about our past. Then, if the content represented in the episodic construction aligns with these propositional beliefs, the *feeling of remembering* will emerge and accompany the episodic representation.[50]

> **c)** ***Comparison with propositional beliefs:*** The feeling of remembering emerges after comparing the episodically represented contents with positional beliefs about our past.

However, this candidate presents two significant drawbacks for STM. First, it heavily undermines the characteristic immediacy of episodic memory. Second, it makes episodic memory dependent on propositional memory and undermines its authority over propositional belief. There are two things to be said about these consequences. On the one hand, Simulationism posits episodic memory is a radically generative source of knowledge. Furthermore, it is crucial to remark that an episodic memory often corrects the content of our propositional beliefs. My belief that

[50] It is important to note that the present argument also works one wants to obviate phenomenology. It could be said that this process of comparing episodically represented content with beliefs is the mechanism for determining whether we are dealing with an episodic memory, regardless of whether this comparative process results in a phenomenological output or a mere judgement.

William did not come to the last seminar might be corrected by the sudden episodic memory of him sitting at the end of the room: "He was there!", I might claim, correcting my previous doxastic state and giving episodic memory authority over belief. On the other hand, experience shows that often we do not have the relevant set of propositional beliefs to compare to and determine the status of the represented episodic content. See, for instance, the following case. Suppose that after looking for your keys around the house for a while, you try to visualize what you did last night when you got home. In trying to remember, the following images come to *the mind's eye* accompanied by the feeling of remembering: the keys falling from the pocket of your coat to the living room floor. They fell on the ground when you left the coat on the chair; you saw them, but tiredness made you postpone bending over to grab them. Based on this sudden image, you form the belief that the keys are on the living room floor. In this case, it seems absurd to say that the identification of the representation as a memory is dependent on a checking process with propositional beliefs, because before the imagery was entertained you did not have any belief about the keys' whereabouts. Therefore, it is not plausible to claim that in such cases, we take episodic constructions as memories after checking its contents with our propositional beliefs about the past. It seems that what makes us endorse them as memories is a much more immediate process.

One possible way to avoid this problem is to suggest that it is not that we compare the represented episodic contents with our propositional beliefs to determine their status, but that the episodic memories are supported by relevant beliefs "about" the contents represented in them. In this line, Debus (2018) claims that episodic memories are "embedded" in a context of relevant beliefs, something that "other" imaginings lack and that let us differentiate between both faculties [51]. Other recent accounts (Redshaw, 2014; Mahr and Csibra, 2018) have also related the phenomenology of

[51] Debus's account is very different from Michaelian's in its metaphysical and phenomenological claims. I bring it up for debate because of the role it attributes to beliefs in determining from a first-personal perspective whether we are remembering.

remembering with the ability to place the representations in a more general narrative of our past. STM could also appeal to this mechanism:

> **d)** ***Embedding in beliefs:*** What causes the feeling of remembering that accompanies episodic memories are propositional beliefs supporting the contents represented in the episodic construction.

This option does not seem to be available for STM either, for two reasons. First, this mechanism inherits the shortcomings of its predecessor. It does not explain paradigmatic cases of memory's authority over belief (like the one of William at the seminar). Since many times episodic memory "corrects" propositional beliefs, it does not seem that its immediacy and authority are due to being surrounded by a set of propositional beliefs. Second, many imaginings are also "embedded" in the context of beliefs about actual states of affairs; namely, they are also "scaffolded" by beliefs about our past. For instance, if a subject entertains an episodic counterfactual thought about what she could have said in an interview after doing it, the imagining will also be constrained and embedded in the context of many beliefs (e.g., beliefs about the interviewer, the management of her emotions, the room where the interview took place, et cetera). This is also the case of many anticipatory imaginings: in the attempt to accurately represent the future, we use many beliefs, as well. Nonetheless, although these episodic imaginings are "embedded" in a context of relevant beliefs and cohere with them, the feeling of remembering does not accompany them.

In this section, we have seen that appealing to propositional beliefs about the past to explain phenomenology is not an apt strategy for the simulationist. If so, it would heavily undermine the immediacy and authority of episodic memory, something the simulationist—who sees episodic memory as a radically generative epistemic source—does not seem to be willing to give up. However, one last candidate remains to which the simulationist can appeal.

4.5.3. Intentions and the aim of the system

Finally, I will consider a last possible candidate for the mechanism underlying the phenomenology of memory: the detection of the aim of the episodic construction system. In STM, the system that produces episodic memories is the same that produces many other episodic imaginings. Nevertheless, the system's aim in doing so is different. In the case of memory, the episodic system aims "to produce a representation of an episode belonging to *S*'s personal past". Thus, a candidate that the simulationist could appeal to for generating the feeling of remembering would be the personal or subpersonal detection of this aim, which is exclusive to memory.

However, what is it for the episodic system to have an *aim* and how could we detect it? It could be the case that the system's aim is the subpersonal dimension of the subject's intention at the personal level. In this case, we could detect the aim of the system by detecting our intentions at the personal level (to remember vs. to imagine). Then, the detection of our intention to remember could bring to the episodic construction the feeling of remembering. Something along these lines was proposed by Urmson (1967)[52]. According to him, we determine whether we are remembering or imagining in the same way in which we would determine who the subject of a portrait is when making a painting: checking our intentions is enough to know what we are doing. In Urmson's words:

> "Let us suppose that a child is drawing what is recognizably a human figure. Let us suppose also, for the sake of definiteness, that the drawing looks quite like Winston Churchill and that the child has as a matter of fact seen Churchill. Now how does the child know whether he is drawing (a) just 'a man', nobody in particular, the resemblance to Churchill being coincidental, or (b) Churchill, or (c)

[52] Urmson dispenses with phenomenological considerations; in his case, the first-personal marker of memory is strictly formal.

> his father (say)? I answer that he has merely to have decided what, if anything, will count for him as success or failure in his enterprise." (1967, p. 86-87)

According to Urmson's criterion, we come to believe that we are remembering by "knowing whether we have or have not chosen to act so that resemblance to actuality is a criterion of the success of our activity" (Urmson 1967: 90). In his view, detecting the criteria of success under which we are entertaining mental images is the crucial element in judging whether we are remembering. This mechanism could be initially endorsed by STM:

- ***Detection of intentions - criteria of success:*** The detection of our own intention to remember and the criteria of success we have established for our activity give rise to the feeling of remembering (or to the judgement that we are remembering).

There are two objections to Urmson's mechanism, which also apply to the potential adoption of this criterion by STM. First, the detection of our intention to remember does not account for the case of unbidden memories: memories that come to our mind without having the intention to remember. Sometimes a memory comes to us, and we recognize it as such without having previously had the intention to remember. Therefore, intention cannot be in these cases—which are common and abundant—the cause of the *feeling of remembering*.

On the other hand, Urmson has been criticized for mistaking *remembering* with *trying to remember* (Teroni, 2017, p.28). It is the case that we can recognize what we are *trying to do* based on our intentions, but this is not enough for us to believe that we are in fact *doing it*. Often, we intend to remember something, and we establish as the rules of recollection criteria for success. Nevertheless, the images that come to mind in an attempt to remember do not satisfy us, so we do not take ourselves to be remembering. In these cases, despite having and recognizing in ourselves the intention to remember, we identify the episodic constructions as imaginings that come to mind. In an analogy: it is not the case that by having the intention of finding gold and carrying a gold detector, I will determine that everything I find is gold. At most, I will determine that I am trying to find gold. Intention does not seem to be an adequate candidate for the simulationist to account for our ability to distinguish remembering from freely imagining.

In the previous sections, we have discarded procedural features and doxastic coherence as mechanisms compatible with STM. Here, we have also discarded intentionality and the system's aims as possible mechanisms. After objecting to the way Simulationism explains the subjective dimension of episodic memory and having explored the more obvious alternative candidates for playing such role, none of them appear compatible with STM. The mechanism underlying the distinctive phenomenology of memory remains ungrounded in the Simulationist paradigm and, with it, the reliability of memory as stated in the theory.

4.6. Conclusion

Any theory of episodic memory must account for our ability to recognize episodic memories as such and distinguish them from imaginative episodes. This requirement is even more significant in STM, which claims that to remember an episode is to simulate it in the imagination. In the present chapter, I have raised two objections to how the first-person recognition of memories is described in the theory. I have shown that the *feeling of remembering* as proposed by Michaelian begs the question of whether STM ground a phenomenological difference between episodic memories and "other" imaginings and leads to incorrect predictions. Here, I have examined potential candidates for the mechanism that allows us to distinguish episodic memories from imaginings. All of them have proven to be either implausible or incompatible with some central claim of Simulationism. In the absence of an explanation of how we distinguish episodic memory from other imaginative episodes in the first-personal, the reliability of memory remains ungrounded in STM. This, in turn, heavily undermines the explanatory power of Simulationism and puts its central ontological assumption under question—namely, that to remember is to imagine. Future development on this issue may concern the revision of some central claims of STM such as the rejection of the Previous Experience Condition or radical generativism about episodic memory.

Conclusions

In this thesis, I have explored several philosophical debates on the role and scope of imagination. I have put forth a view on the architecture and functional role of experiential imagination—the *Prima Facie* View—and a new account of intrinsic symbolic actions. With a more critical aim, I have raised issues regarding appealing to the imagination to account for the functional profile of delusions and the constructive character of episodic memory. I will now summarize what I have done in the previous chapters and the future lines of research each points to.

In Chapter 1, I proposed a view on the architecture and functional profile of experiential imagination: the *Prima Facie* View. I accounted for clinical and empirical evidence on the effects of experiential imagination on attitudes, emotions, and behavior. According to the *Prima Facie* View, in several subpersonal operations, the quasi-sensory contents of experiential imaginings are processed at face value (i.e., *prima facie*). I have motivated the claim that given its *implicit assertoric force*—a notion I have coined—imagination is not an epistemically innocuous enterprise.

The connections between the *Prima Facie* View and Spinozian models of belief formation open up the door for further research on this topic. According to Spiniozian models, the mere activation of a truth-apt proposition leads to believing it. To my knowledge, no partisan of Spinozianism has mentioned differential effects on belief depending on the format in which the truth-apt proposition is entertained. However, the evidence presented here shows that the imagistic representation format (and its vivacity) is relevant in mediating the effects seen on attitudes and behavior. It is to be seen whether a Spinozian model of belief that considers the format in which representations are entertained—imagistic versus purely propositional—can give an account of the evidence reviewed in Chapter 1. Another research avenue this dissertation opens up concerns the role of mental imagery in the associative system as well as in the so-called intuitive system in dual-system theory. One plausible explanation for this phenomenon is that the associative system gives more relevance to the format (imagistic vs. propositional) than to the source (internal—imagination—versus external—perception). I hypothesize that, given its faster and associative nature, the intuitive system might prioritize the imagistic representations in its heuristics since it is how first-hand evidence (i.e., perception and episodic memories) is usually formatted.

I have opened Chapter 2 by giving a direct definition of symbolic actions. After distinguishing between instrumental and intrinsic symbolic actions, I have put forth three desiderata for accounts of why agents perform these actions. I then offered a new account of symbolic actions. By this account, when performing a symbolic action, the agent simultaneously displaces an emotional action tendency and imagines himself doing a thwarted action. The object of the action, in turn, stands in a symbolic relationship with the absent object represented in imagining, providing symbolic satisfaction. I offered clarification on both the notion of symbolic meaning and the conditions for an object to stand in a symbolic relationship with another. Further research is necessary regarding the nature of symbolic satisfaction and the phenomenon of displacement. Can thwarted action tendencies be displaced unconsciously? What function does displacement serve in the emotional economy of the subject? What is the legitimacy of such claims in psychoanalytical practice? Dissertations on symbolism and symbolic actions have been the patrimony of continental dissertations.

However, an analytic account of the concepts involved and the legitimacy of the assumptions of psychoanalytical practice is necessary.

In short, in the first part of this thesis, I have accounted for the consequences that imagining has on our conception of the world and our actions. As I have pointed out, imagining experiences our self-concept or perception of events influences in insidious and unnoticed ways, which has implications on many levels. One of the most seemingly remote is the political domain. The continuing contemporary appeal to imagine and dream is countered in this dissertation. Imagining, for example, can lead to a loss of motivation in the consummation of goals. Due to its capacity to generate emotions, imagination may serve in the short term to cope with frustration and provide satisfaction. However, in the long run, immersing oneself excessively in it might be counter-productive—even in non-pathological subjects and when imagination is properly monitored.

While in the first part of the thesis, I emphasized the effects of imagination that have not been sufficiently considered, in the second part I delimited the appeal to the imagination in two debates: the delusion and the episodic memory debates.

In Chapter 3, I have raised criticisms of the account that takes delusions to match the functional profile of imaginings. I have then modeled delusions in a fragmented and psychofunctional system of belief. Further research on the nature of delusions in a fragmented model concerns how patients deal with the coactivation of fragments containing contradictory beliefs (the delusional and the realistic). Given that the delusional belief is context-sensitive, further investigation on the conditions needed to access the delusional belief or its opposite is needed. One crucial aspect is the need for the presence of an *abnormal experience*—or its memory—as a limiting factor in accessing the delusional beliefs. The idea of the psychological immune system as a mechanism mediating belief updating needs further development, especially since we often tend to irrationally endorse beliefs that go against positive core beliefs about ourselves. Or is it that we only *seem* to do so?

In Chapter 4, I raised criticisms about subsuming episodic memory in imagination in the way proposed by the Simulation Theory of memory. I have raised two objections to how the feeling of remembering is described in the Theory, followed by an exhaustive exploration of the Theory's ability to ground the mechanism underlying this feeling. I have concluded that the Simulation Theory cannot simultaneously defend the simulational character of episodic memory and ground our ability to discriminate between memories and imaginings. It is worth pursuing further research concerning the episodic constructive system endorsed by simulationists. This system is postulated as responsible for simulating a wide variety of episodes—from episodic anticipation to episodic memory. As stated in Chapter 4, the system allows for the conscious incorporation of information when anticipating the future but does not allow it when remembering past events. This input asymmetry is problematic for the individuation of the system and needs to be addressed by Simulationists. Regarding the need to ground the phenomenology of memory, amendments in the Theory—such as embracing the Previous Experience Condition—could help postulate a procedural feature underlying the phenomenology of memory.

The final extension and implication of imagination in our mental life is still not fully understood. This thesis attempts to be a small step forward. The use of our highest abilities, among them imagination, will be required in the future to solve its mysteries.

References

Abed, R. T., and Fewtrell, W. D. (1990). Delusional misidentification of familiar inanimate objects. *British Journal of Psychiatry, 157*, 915–917.

Abramowitz, J. S. (1996). Variants of exposure and response prevention in the treatment of obsessive-compulsive disorder: A meta-analysis. *Behavior therapy*, *27*(4), 583-600.

Addis, D., A.A.T. Wong and D. L. Schacter. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia* 45 (7): 1363-1377.

Alexander, M. P., Stuss, D. T. and Benson, D. F. (1979). Capgras Syndrome: A Reduplicative Phenomenon, *Neurology* 29(3), pp. 334–339.

Allport, G. W., Clark, K., & Pettigrew, T. (1954). *The nature of prejudice.* Cambridge MA: Perseus Books.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of MentalDisorders.* Fifth Edition. Washington, DC: American Psychiatric Association.

Anderson, C. A. (1983). Imagination and expectation: The effect of imagining behavioral scripts on personal influences. *Journal of personality and social psychology*, *45*(2), 293.

Antony, M.M., & Swinson, R.P. (2000). Specific phobia. In: Antony, M.M., Swinson, R.P. (Eds.), *Phobic disorders and Panic in Adults: A Guide to Assessment and Treatment.* American Psychological Association: Washington, DC. pp. 79–104.

Arntz, A., Tiesema, M., & Kindt, M. (2007). Treatment of PTSD: A comparison of imaginal exposure with and without imagery rescripting. *Journal of behavior therapy and experimental psychiatry*, *38*(4), 345-370.

Asbrock, F., Gutenbrunner, L., & Wagner, U. (2013). Unwilling, but not unaffected—Imagined contact effects for authoritarians and social dominators. *European Journal of Social Psychology*, *43*(5), 404-412.

Audi, R. (1995). Memorial justification. *Philosophical Topics* 23: 31-45.

Badura, C., & Kind, A. (Eds.). (2021). *Epistemic uses of imagination*. London: Routledge.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

Bayne, T. and Pacherie, E. (2004). Bottom-up or top-down? Campbell's rationalistaccount of monothematic delusions. *Philosophy, Psychiatry, & Psychology, 11*, 1–11.

Bayne, T. & Pacherie, E. (2005). In defense of the doxastic conception of delusions. *Mind & Language, 20* (2), 163–188.

Bayne. T & Fernández. J (2019), *Delusion and Self-Deception: Affective and Motivational Influences on Belief Formation.* Hove: Psychology Press.

Bendaña, J and Mandelbaum, E. (2021). The fragmentation of belief. In D. Kinderman, A. Onofri, C. Borgoni (eds), *The Fragmented Mind.* Oxford: Oxford University Press.

Benoit, R. G., Paulus, P. C., & Schacter, D. L. (2019). Forming attitudes via neural activity supporting affective episodic simulations. *Nature communications*, *10*(1), 1-11.

Bernecker, S. (2008). *The metaphysics of memory.* Springer.

Bernecker, S. (2010). *Memory: A Philosophical Study.* Oxford: Oxford University Press.

Bleuler, E. (1924). *Textbook of Psychiatry* (4th ed), trans. AA Brill. New York: Macmillan.

Borgoni, C., Kindermann, D., & Onofri, A. (Eds.). (2021). *The Fragmented Mind.* Oxford University Press.

Bortolotti, L. (2010). *Delusions and other irrational beliefs.* Oxford: Oxford UniversityPress.

Bortolotti, L. (2011). Double bookkeeping in delusions: Explaining the gap between saying and doing. In *New Waves in Philosophy of Action* (pp. 237-256). Palgrave Macmillan, London.

Bortolotti, L., & Mameli, M. (2012). Self-deception, delusion and the boundaries of folk psychology. *Humanamente*, *20*, 203.

Bortolotti, L. (2018) Delusion. *The Stanford Encyclopedia of Philosophy.* Edward N. Zalta (ed.).

Botella, C., Fernández-Álvarez, J., Guillén, V., García-Palacios, A., & Baños, R. (2017). Recent progress in virtual reality exposure therapy for phobias: a systematic review. *Current psychiatry reports*, *19*(7), 1-13.

Brainerd, C. J., and V. F. Reyna. (2005). *The Science of False Memory.* Oxford: Oxford University Press.

Bryant, R. A., Moulds, M. L., Guthrie, R. M., Dang, S. T., & Nixon, R. D. (2003). Imaginal exposure alone and imaginal exposure with cognitive restructuring in treatment of posttraumatic stress disorder. *Journal of consulting and clinical psychology*, *71*(4), 706.

Bullier, J. (2001). Integrated model of visual processing. Brain Research Reviews, 36, 96–107.

Bullier, J. (2004). Communications between cortical areas of the visual system. In L. M. Chalupa & J. S. Werner (Eds.), The visual neurosciences (pp. 522–540). MIT Press.

Byrne, A. (2010). Recollection, Perception, Imagination. *Philosophical Studies*, 148(1): 15–26. doi:10.1007/s11098-010-9508-1.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., ... & Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637-644.

Capgras, J., & Carrette, P. (1924). Illusion des sosies et complexe d'Oedipe. *AnnalesMédico- Psychologiques, 12*, 48–68.

Carroll, J. S. (1978). The effect of imagining an event on expectations for the event: An interpretation in terms of the availability heuristic. *Journal of experimental social psychology*, *14*(1), 88-96.

Cassirer, E. (1944). *An Essay on Man.* Yale University Press.

Chasid, A., & Weksler, A. (2020). Belief-like imaginings and perceptual (non-) assertoricity. *Philosophical Psychology*, *33*(5), 731-751.

Choy Y, Fyer AJ, Lipsitz JD. (2007) Treatment of specific phobia in adults. *Clin Psychol Rev* ; 27: 266–86.

Cohen, J., & Weimann, G. (2000). Cultivation revisited: Some genres have some effects on some viewers. *Communication reports*, *13*(2), 99-114

Coltheart, M. (2005). Conscious Experience and Delusional Belief. *Philosophy,Psychiatry and Psychology, 12*, pp. 153–7.

Coltheart, M. (2007). The 33rd Sir Frederick Bartlett Lecture. Cognitive neuropsychiatry and delusional belief. *Quarterly Journal of Experimental Psychology. 60* (8):1041-1062.

Coltheart, M., Menzies, P. and Sutton, J. (2010). Abductive Inference and DelusionalBelief. *Cognitive Neuropsychiatry, 15*, pp. 261–87.

Coltheart, M., Langdon, R. and McKay, R. (2011). Delusional Belief. *Annual Reviewof Psychology, 62*, pp. 271–98.

Coltheart, M., & Davies, M. (2022). What is Capgras delusion? *Cognitive Neuropsychiatry*, *27*(1), 69-82.

Crisp, R. J., & Turner, R. N. (2009). Can imagined interactions produce positive perceptions?: Reducing prejudice through simulated social contact. *American psychologist*, *64*(4), 231.

Currie, G. (2000). Imagination, delusion, and hallucinations. In Coltheart, M. and Davies,M (eds.), *Pathologies of Belief.* Blackwell, 167–182.

Currie, G. and Jureidini, J. (2001). Delusion, rationality, empathy. *Philosophy, Psychiatry& Psychology, 8*/2,3, 159–62.

Currie, G. and Ravenscroft, I. (2002). *Recreative Minds.* Oxford: Oxford UniversityPress.

Dadds, M. R., Bovbjerg, D. H., Redd, W. H., & Cutmore, T. R. (1997). Imagery in human classical conditioning. *Psychological bulletin*, *122*(1), 89.

Davidson, D. (1963) Actions, reasons and causes. *The Journal of Philosophy, Vol. 60*, No. 23, American Philosophical Association, Eastern Division, Sixtieth Annual Meeting. pp. 685-700

Davies, M., Coltheart, M., Langdon, R. and Breen, N. (2001). Monothematic Delusions: Towards a Two-Factor Account. *Philosophy, Psychiatry and Psychology, 8*, pp. 133–58.

Davies, M., and Egan, A. (2013). Delusion: Cognitive Approaches, Bayesian Inference,and Compartmentalization. In K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Sadler, G. Stanghellini and T. Thornton (eds), *The Oxford Handbook of Philosophy ofPsychiatry*. Oxford: Oxford University Press.

De Brigard, F. (2014a). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese* 191 (2): 155-185

De Brigard, F. (2014b). The nature of memory traces. *Philos. Comp.* 9, 402–414. doi: 10.1111/phc3.12133.

De Brigard, F. (2017). Memory and Imagination, in Bernecker & Michaelian (eds.), *The Routledge Handbook of Philosophy of Memory*. London: Routledge. Chapter 10: 127-140.

de Pauw, K.W., and T. K. Szulecka. (1988). Dangerous delusions: Violence and themisidentification syndromes. *British Journal of Psychiatry 152*: 91–97.

De Sousa, R. (2007) Emotion. *The Stanford Encyclopedia of Philosophy.* (Winter 2007 Edition)

https://stanford.library.sydney.edu.au/archives/win2007/entries/emotion/

Debus, D. (2018). Memory, imagination, and narrative. In Macpherson, F & Dorsch, F, *Perceptual Imagination and Perceptual Memory.* Oxford University Press: Oxford.

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. (2019). Shared neural mechanisms of visual perception and imagery. *Trends in cognitive sciences*, *23*(5), 423-434.

Dijkstra, N., Kok, P., & Fleming, S. M. (2022). Perceptual reality monitoring: Neural mechanisms dissociating imagination from reality. *Neuroscience & Biobehavioral Reviews*, 104557.

Dokic, J. (2014). Feeling the past: A two-tiered account of episodic memory. *Review of Philosophy and Psychology.* 5 (3): 413-42

Dolcos, S., & Albarracin, D. (2014). The inner speech of behavioral regulation: Intentions and task performance strengthen when you talk to yourself as a You. *European Journal of Social Psychology*, *44*(6), 636-642.

Dummett, M. (1994). Memory and testimony. In *Knowing from Words: Western and Indian Philosophical Analysis of Understanding and Testimony,* ed. B.K. Matilal and A. Chakrabarty, 251-272. Springer.

Eaton, W. W., Bienvenu, O. J., & Miloyan, B. (2018). Specific phobias. *The Lancet Psychiatry*, *5*(8), 678-686.

Egan, A. (2008). Seeing and Believing: Perception, Belief Formation, and the DividedMind. *Philosophical Studies 140.1*: 47–63.

Egan, A. (2021) Fragmented Models of Belief. In Kindermann, D., Onofri, A.,Borgoni, C. *The fragmented mind.* Oxford: Oxford University Press.

Ellis, HD and Young, AW. (1990). Accounting for delusional misidentifications. *Br.J.Psychiatry 157*:239–48.

Ellis HD, Young AW, Quayle AH, de Pauw KW. (1997). Reduced autonomic responses to faces in Capgras delusion. *Proc.R.Soc.Ser.BBiol.Sci. 264*: 1085–92.

Farah, M. J. (1984). The neurological basis of mental imagery: A componential analysis. *Cognition*, *18*(1-3), 245-272.

Fernàndez, J. (2019). *Memory: A self-referential account.* Oxford University Press: Oxford.

Field, A. P. (2006). Is conditioning a useful framework for understanding the development and treatment of phobias?. *Clinical Psychology Review*, *26*(7), 857-875.

Foa, E. B., Steketee, G., Turner, R. M., & Fischer, S. C. (1980). Effects of imaginal exposure to feared disasters in obsessive-compulsive checkers. *Behaviour Research and Therapy*, *18*(5), 449-455.

Fazio, L. K., Perfors, A., & Ecker, U. (2020). Repetition increases perceived truth even for known falsehoods. *Collabra: Psychology*, *6*(1).

Förstl, H., O. P. Almeida, A. M. Owen, A. Burns, and R. Howard. (1991). Psychiatric, neurological and medical aspects of misidentification syndromes: A review of 260 cases.*Psychological Medicine 21*: 905–910.

Freud, S. (1914) *The Interpretation of Dreams*, Standard Edition, vols. IV and V, London: Hogarth.

Freud, S. (1916) *A connection between a symbol and a symptom*. Standard Edition, vol. XIV, London: Hogarth.

Freud, S. (1917) *Introductory Lectures on Psycho-Analysis*, Standard Edition, vols.XV and XVI, London: Hogarth.

Garry, M. et al. (1996) Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychon. Bull. Rev.* 3, 208–214.

Gendler, T. S. (2013). *Intuition, imagination, and philosophical methodology*. OUP Oxford.

Gilbert, D. T. (1991). How mental systems believe. *American psychologist*, *46*(2), 107.

Goff, L.M. and Roediger, H.L., III (1998) Imagination inflation for action events: Repeated imaginings lead to illusory recollections. *Mem. Cognit.* 26, 20–33.

Goldie, P. (2000). Explaining expressions of emotion. *Mind*, *109*(433), 25-38.

Grayson, J. B. (1982). The elicitation and habituation of orienting and defensive responses to phobic imagery and the incremental stimulus intensity effect. *Psychophysiology*, *19*(1), 104-111.

Gregory, W. L., Cialdini, R. B., & Carpenter, K. M. (1982). Self-relevant scenarios as mediators of likelihood estimates and compliance: Does imagining make it so? *Journal of Personality and Social Psychology*, 43(1), 89.

Hackmann, A., Bennett-Levy, J., & Holmes, E. A. (2011). *Oxford guide to imagery in cognitive therapy*. Oxford university press.Halligan, P.W. and Marshall, J.C. (1995). Supernumerary phantom limb after right hemisphere stroke. *Journal of Neurology, Neurosurgery and Psychiatry, 59*, 341–2.

Hanrahan, R. (2021). Crossing Rivers: Imagination and Real Possibilities. In *Epistemic Uses of Imagination* (pp. 63-78). Routledge.

Hardy, J., Thomas, A. V., & Blanchfield, A. W. (2019). To me, to you: How you say things matters for endurance performance. *Journal of Sports Sciences*, *37*(18), 2122-2130.

Hirstein, WS and Ramachandran, VS. (1997). Capgras syndrome: a novel probe for understanding the neural representation of the identity and familiarity of persons. Proc.R. Soc. Lond. B 264:437–44.

Hoerl, C. (2001). The Phenomenology of episodic Recall. In C. Hoerl and T. McCormark (eds.) *Time and Memory*. Oxford: Oxford University Press.

Hoerl, C. (2019). A 'knowledge-first' approach to episodic memory. In *Proceedings of the Keynote talk delivered at "Issues in Philosophy of Memory 2",* Grenoble.

Hopkins, R. (2018). Imagining the past: On the nature of episodic memory. In Macpherson, F & Dorsch, F, *Perceptual Imagination and Perceptual Memory*. Oxford University Press: Oxford.

Hursthouse, R. (1991). Arational actions. *The Journal of Philosophy*, *88*(2), 57-68

Husnu, S., & Crisp, R. J. (2011). Enhancing the imagined contact effect. *The Journal of social psychology*, *151*(1), 113-116.

Islam L, Piacentini S, Soliveri P, Scarone S, Gambini O. (2015) Capgras delusion for animals and inanimate objects in Parkinson's Disease: a case report. *BMC Psychiatry; 15*: 73.

James, W. (1890). *The Principles of Psychology.* 1st volume. London: Macmillan.

Kahneman, D., & Tversky, A. (1981). *The simulation heuristic.* Stanford University. CA: Deptartment of Psychology.

Kappes, H. B., & Oettingen, G. (2011). Positive fantasies about idealized futures sap energy. *Journal of Experimental Social Psychology*, 47(4), 719–729.

Kappes, H. B., Oettingen, G., & Mayer, D. (2012). Positive fantasies predict low academic achievement in disadvantaged students. *European Journal of Social Psychology*, 42(1), 53–64.

Kappes, H. B., Schwörer, B., & Oettingen, G. (2012). Needs instigate positive fantasies of idealized futures. *European Journal of Social Psychology*, 42,299–307.

Kappes, H. B., & Morewedge, C. K. (2016). Mental simulation as substitute for experience. *Social and Personality Psychology Compass*, *10*(7), 405-420.

Koriat A. (2007). Metacognition and consciousness In *The Cambridge Handbook of Consciousness*, eds Zelazo P. D., Moscovitch M., Thompson E. New York: Cambridge University Press.

Kosslyn, S. M. (1980). *Image and Mind.* Cambridge, MA: Harvard University Press.

Kosslyn, S. M. (1994). *Image and Brain:* The Resolution of the Imagery Debate.

Cambridge, MA: The MIT Press.

Kosslyn, S. M., & Shin, L. M. (1991). Visual mental images in the brain. *Proceedings of the American Philosophical Society*, *135*(4), 524-532.

Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological bulletin*, *129*(5), 723.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature reviews neuroscience*, *2*(9), 635-642.

Kovach, A., & De Lancey, C. (2005). On emotions and the explanation of behavior. *Nous*, *39*(1), 106-122.

Kross, E., Bruehlman-Senecal, E., Park, J., Burson, A., Dougherty, A., Shablack, H., ... & Ayduk, O. (2014). Self-talk as a regulatory mechanism: how you do it matters. *Journal of personality and social psychology*, *106*(2), 304.

Kruger, J., & Gilovich, T. (2004). Actions, intentions, and self-assessment: The road to self-enhancement is paved with good intentions. *Personality and Social Psychology Bulletin*, *30*(3), 328-339.

Kung, P. (2010). Imagining as a guide to possibility. *Philosophy and Phenomenological Research*, *81*(3), 620-663.

Langdon, R., Connaughton, E., & Coltheart, M. (2014). The Fregoli delusion: a disorder of person identification and tracking. *Topics in cognitive science*, *6*(4), 615-631.

Langer, S. K. (1942 .2009). *Philosophy in a new key: A study in the symbolism of reason, rite, and art.* Harvard University Press.

Langland-Hassan, P. 2012. "Pretense, Imagination, and Belief: The Single Attitude Theory." *Philosophical Studies.* 159 (2): 155–179.

Lewis, D. E., O'Reilly, M. J., Khuu, S. K., & Pearson, J. (2013). Conditioning the mind's eye: Associative learning with voluntary mental imagery. *Clinical Psychological Science*, *1*(4), 390-400.

Lewis, M. (2016). *The undoing project: A friendship that changed the world.* Penguin UK.

Libby, L. K., Shaeffer, E. M., Eibach, R. P., & Slemmer, J. A. (2007). Picture yourself at the polls: Visual perspective in mental imagery affects self-perception and behavior. *Psychological Science*, *18*(3), 199-203.

Loftus, E.F. (2003) Make-believe memories. *Am. Psychol.* 58, 867–873.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory* 12 (4): 361-366.

Lorenz, K. (1957) The past twelve years in the com- parative study of behavior. In G. H. Schiller (Ed.), *Instinctive behavior.* New York: International Univer. Press, pp. 288-310.

Lucchelli, F., and H. Spinnler. (2007). The case of lost Wilma: a clinical report of Capgrasdelusion. *Neurological Science 28*(4): 188–195.

Macpherson, F., and Dorsch, F. (Eds.). (2018). *Perceptual Imagination and Perceptual Memory.* Oxford: Oxford University Press.

Maher, B. (1974). Delusional thinking and perceptual disorder. *Journal of Individual Psychology, 30*, 98–113.

Maher, B.A., 1999. Anomalous experience in everyday life: Its significance for psychopathology, *The Monist*, 82: 547–70.

Mahr, J.B. & Csibra, G. (2018). Why do we remember? The communicative function of episodic memory, *Behavioral and Brain Sciences, 41*, e1.

Mahr, J. (2020). The dimensions of episodic simulation. *Cognition.* 196: 104085 doi: 10.1016/j.cognition.2019.104085.

Mandelbaum, E. (2013). Against alief. *Philosophical studies*, *165*(1), 197-211.

Mandelbaum, E. (2014). Thinking is believing. *Inquiry*, *57*(1), 55-96.

Mandelbaum, E. (2019). Troubles with Bayesianism: An introduction to the psychological immune system. *Mind and Language 34* (2): 141-157.

Mandelbaum, E. (2020) Associationist Theories of Thought, in *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta (ed.), URL:

https://plato.stanford.edu/archives/fall2020/entries/associationist-thought/

Martin, C. B., and M. Deutscher. (1966). Remembering. *Philosophical Review* 75: 161-196.

Martin, M.G.F. (2001). Out of the past: Episodic Recall as Retained Acquaintance. In C. Hoerl and T. McCormark (eds.) *Time and Memory.* Oxford: Oxford University Press.

McCarroll, C. J. (2020) Remembering the personal past: Beyond the boundaries of imagination. *Frontiers in Psychology.* Volume 12. Article 585352. 1-10.

McDermott, Drew. (1987). We've been framed: Or, why AI is innocent of the frame problem. In Zenon W. Pylyshyn (ed.), *The Robot's Dilemma.* Westport, CT: Greenwood Publishers.

McKay, R. (2012). Delusional Inference, *Mind and Language, 27*, pp. 330–55.

Meade, M. L, and H. L. Roediger. (2002). Explorations in the social contagion of memory. *Memory & Cognition* 30 (7): 995 – 1009.

Mele, A. (2006). Self-deception and delusions. *European Journal of Analytic Philosophy*, *2*(1), 109-124.

Mertens, G., Krypotos, A. M., & Engelhard, I. M. (2020). A review on mental imagery in fear conditioning research 100 years since the 'Little Albert'study. *Behaviour Research and Therapy*, *126*, 103556.

Michaelian, K. (2016a). *Mental time travel: Episodic memory and our knowledge of the personal past.* Cambridge: MIT Press.

Michaelian, K. (2016b). Against discontinuism: Mental time travel and our knowledge of past and future events. In K. Michaelian, S. Klein, and K. Szpunar (eds) *Seeing the Future: Theoretical Perspectives of Future Oriented Mental Time Travel.* pp. 62-92).

Michaelian, K. (2021). Imagining the past reliably and unreliably: Towards a virtue theory of memory. *Synthese* 199 (3-4): 7477-7507. Special issue: *Imagination and Its Limits.* Eds. Amy Kind and Tufan Kıymaz.

Miles, E., & Crisp, R. J. (2014). A meta-analytic test of the imagined contact hypothesis. *Group Processes & Intergroup Relations*, *17*(1), 3-26.

Mill, J. S. (1963). System of logic: Ratiocinative and inductive. In *Collected works of John Stuart Mill.* Volume VIII. Toronto: University of Toronto Press. (Original work published 1843)

Minnen, A. V., & Foa, E. B. (2006). The effect of imaginal exposure length on outcome of treatment for PTSD. *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies*, *19*(4), 427-438.

Morewedge, C. K., Huh, Y. E., & Vosgerau, J. (2010). Thought for food: Imagined consumption reduces actual consumption. *Science*, *330*(6010), 1530-1533.

Morris, A., Gaesser, B., & Cushman, F. (2022). The role of episodic simulation in motivating commonplace harms. *Cognition*, *225*, 105104.

Moser, J. S., Dougherty, A., Mattson, W. I., Katz, B., Moran, T. P., Guevarra, D., ... & Kross, E. (2017). Third-person self-talk facilitates emotion regulation without engaging cognitive control: Converging evidence from ERP and fMRI. *Scientific reports*, *7*(1), 1-9.

Mueller, E. M., Sperl, M. F., & Panitz, C. (2019). Aversive imagery causes de novo fear conditioning. *Psychological Science*, *30*(7), 1001-1015.

Mullally, S. L., & Maguire, E. A. (2014). Memory, Imagination, and Predicting the Future: A Common Brain Mechanism? *The Neuroscientist,* 20 (3), 220–234. https://doi.org/10.1177/1073858413495091

Nanay, B. (2021). Imagining one experience to be another. *Synthese*, *199*(5), 13977-13991.

Nanay, B. (2015). Perceptual content and the content of mental imagery. *Philosophical Studies*, *172*(7), 1723-1736.

Nichols, S. 2004. Imagining and Believing: The Promise of a Single Code. *Journal of Aesthetics and Art Criticism.* 62 (2): 129–139.

Oettingen, G., & Mayer, D. (2002). The motivating function of thinking about the future: Expectations versus fantasies. *Journal of Personality and Social Psychology*, 83(5), 1198.

Oettingen, G., & Wadden, T. A. (1991). Expectation, fantasy, and weight loss: Is the impact of positive thinking always positive? *Cognitive Therapy and Research*, 15(2), 167–175.

Öst, L. G., Svensson, L., Hellström, K., & Lindwall, R. (2001). One-Session treatment of specific phobias in youths: a randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *69*(5), 814.

Pacherie, E., Green, M. and Bayne, T., 2006. Phenomenology and delusions: Who put the 'alien' in alien control? *Consciousness and Cognition*, 15: 566–577.

Pacherie, E. (2009). Perception, Emotions, and Delusions. The Case of the Capgras Delusion. In Bayne. T & Fernández. J (eds.), *Delusion and Self-Deception: Affective andMotivational Influences on Belief Formation.* Hove: Psychology Press

Pandis, C., Agrawal, N., Poole, N. (2019). Capgras' Delusion: A Systematic Review of 255 Published Cases. *Psychopathology. 52* (3): 161-173.

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: functional mechanisms and clinical applications. *Trends in cognitive sciences*, *19*(10), 590-602.

Pearson, J. (2019). The human imagination: the cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience*, *20*(10), 624-634.

Perrin, D., Michaelian, K and Sant'Anna, A. (2020). The Phenomenology of Remembering Is an Epistemic Feeling. *Front. Psychol. 11*: 1531. doi: 10.3389/fpsyg.2020.01531

Petocz, A. (1999). *Freud, psychoanalysis and symbolism.* Cambridge University Press.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of personality and social psychology*, *90*(5), 751.

Petrova, P. K., & Cialdini, R. B. (2008). Evoking the imagination as a strategy of influence. In C. P. Haugtvedt, P. M. Herr, & F. R. Kardes (Eds.), *Handbook of consumer psychology* (pp. 505–523). Taylor & Francis Group/Lawrence Erlbaum Associates.

Poupart, F., Bouscail, M., Sturm, G., Bensoussan, A., Galliot, G., & Gozé, T. (2021). Acting on delusion and delusional inconsequentiality: A review. *Comprehensive psychiatry*, *106*, 152230.

Quilty-Dunn, J. & Mandelbaum, E. (2017). Against dispositionalism: Belief in cognitivescience. *Philosophical Studies ,1–20*

Rachman, S. (1967). Systematic desensitization. *Psychological Bulletin*, 67(2), 93.

Radden, J. (2014). Belief as delusional and delusion as belief. *Philosophy, Psychiatry, &Psychology, 21*(1), 43–46.

Redshaw, J. (2014). Does metarepresentation make human mental time travel unique? Wiley Interdisciplinary Reviews: Cognitive Science 5 (5): 519 – 531

Rentz, T. O., Powers, M. B., Smits, J. A., Cougle, J. R., & Telch, M. J. (2003). Active-imaginal exposure: Examination of a new behavioral treatment for cynophobia (dog phobia). *Behaviour Research and Therapy*, *41*(11), 1337-1353.

Rescorla, R. A. (1976). Stimulus generalization: Some predictions from a model of Pavlovian conditioning. *Journal of Experimental Psychology:*

Animal Behavior Processes, 2(1), 88–96. https://doi.org/10.1037/0097-7403.2.1.88

Robins, S. K. (2017). Memory traces. In *The Routledge Handbook of Philosophy of Memory*, eds S. Bernecker and K. Michaelian. London: Routledge p 76–87. doi: 10.4324/9781315687315-7

Russell, B. (1921). *The Analysis of Mind*. London: Routledge.

Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., & Castiello, U. (2008). How to accurately detect autobiographical events. *Psychological science*, *19*(8), 772-780.

Sass L. (1994). *The Paradoxes of Delusion: Witttgenstein, Schreber and the Schizophrenic Mind*. Ithaca, NY: Cornell University Press.

Scarantino, A., (2014). The motivational theory of emotions. In J. D'arms & D.Jacobson (Eds.), Moral psychology and human agency: Philosophical essays on the science of ethics (pp 156-185). Oxford university press.

Scarantino, A., & Nielsen, M. (2015). Voodoo dolls and angry lions: How emotions explain arational actions. *Philosophical Studies*, *172*(11), 2975-2998.

Sherman, S. J., Cialdini, R. B., Schwartzman, D. F., & Reynolds, K. D. (1985). Imagining can heighten or lower the perceived likelihood of contracting a disease: The mediating effects of ease of imagery. *Personality and Social Psychology Bulletin*, 11, 118-127

Shamloo, S. E., Carnaghi, A., Piccoli, V., Grassi, M., & Bianchi, M. (2018). Imagined intergroup physical contact improves attitudes toward immigrants. *Frontiers in psychology*, 1685.

Schellenberg, S. 2013. "Belief and Desire in Imagination and Immersion." *Journal of Philosophy* 110 (9): 497–517.

Shidlovski, D., Schul, Y., & Mayo, R. (2014). If I imagine it, then it happened: The implicit truth value of imaginary representations. *Cognition*, *133*(3), 517-529.

Skorupski, J. (1976). *Symbol and theory*. Cambridge University Press.

Smith, M. (1998). The possibility of philosophy of action. In *Ethics and the a priori*. (p. 55-180). Cambridge University Press.

Sparing, R., Mottaghy, F. M., Ganis, G., Thompson, W. L., Töpper, R., Kosslyn, S. M., & Pascual-Leone, A. (2002). Visual cortex excitability increases during visual mental imagery—a TMS study in healthy human subjects. *Brain research*, *938*(1-2), 92-97.

Stock, K. (2017). *Only Imagine: Fiction, Interpretation, and Imagination*. Oxford: Oxford University Press

Stone, T., & young, A. (1997). Delusions and brain injury: The philosophy and psychology of belief. *Mind and Language, 12* (3/4), 327–364.

Szpunar, K.K., Watson J.M, and McDermott, K.B. (2007). Neural substrates of envisioning the future. *PNAS* 104: 642-647

Szpunar, K.K. (2010). Episodic future thought: an emerging concept. *Perspect Psychol Sci* 5(2): 142-162.

Szpunar KK, Schacter DL (2013) Get real: Effects of repeated simulation and emotion on the perceived plausibility of future experiences. *J Exp Psychol Gen* 142(2):323–327.

Taber, C. S. & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50* (3), 755–769.

Teroni, F. (2017). The phenomenology of memory. in Bernecker & Michaelian (eds.), *The Routledge Handbook of Philosophy of Memory*, London: Routledge. Chapter 2.

Tinbergen, N. (1939). On the analysis of social organization among vertebrates, with special reference to birds. *American Midland Naturalist, 21*(1), 210–234.

Tinbergen, N. (2020). *The study of instinct.* Pygmalion Press, an imprint of Plunkett Lake Press.

Tulving, E. (1985). Memory and consciousness. *Can. Psychol. Psychol.* 26: 1–12.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology* 53: 1–25.

Turner, R. N., Crisp, R. J., & Lambert, E. (2007). Imagining intergroup contact can improve intergroup attitudes. *Group Processes & Intergroup Relations*, *10*(4), 427-441.

Turner, R. N., & Crisp, R. J. (2010). Imagining intergroup contact reduces implicit prejudice. *British Journal of Social Psychology*, *49*(1), 129-142.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, *5*(2), 207-232.

Vigotsky. L. (1933) Play and its role in the Mental development of the Child. *International Research in Early Childhood Education, 2016, v7* n2 pp. 3-25.

Walton, K. L. (1990). *Mimesis as make-believe: On the foundations of the representational arts.* Harvard University Press.

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of experimental psychology*, *3*(1), 1.

Weinberg, J., & Meskin, A. (2006). Puzzling over the imagination: Philosophical problems, architectural solutions. In S. Nichols (Ed.), *The architecture of the imagination: New essays on pretense, possibility and fiction.* Oxford, UK: Oxford University Press.

Werning, M. (2020). Predicting the past from minimal traces: episodic memory and its distinction from imagination and preservation. *Rev. Philos. Psychol.* 11, 301–333. doi: 10.1007/s13164-020-00471-z.

Whittlesea, B. (1997). Production, evaluation, and preservation of experiences: constructive processing in memory and performance tasks. *Psychol. Learn. Motiv.* 37: 211–264. doi: 10.1016/s0079-7421(08) 60503-4.

Whittlesea, B., and Leboe, J. P. (2000). The heuristic basis of remembering and classification. *J. Exp. Psychol. Gen.* 129: 84–106. doi: 10.1037/0096-3445. 129.1.84.

Wolitzky-Taylor KB, Horowitz JD, Powers MB, Telch MJ. (2008) Psychological approaches in the treatment of specific phobias: a meta-analysis. *Clin Psychol Rev*; 28: 1021–37.

Wolpe, J. (1982). *The practice of behavior therapy* (3rd ed.). New York: Pergamon Press Inc.

Yablo, S. (1993). Is conceivability a guide to possibility? *Philosophy and Phenomenological Research*, *53*(1), 1-42.

Young, A. W., Leafhead, K. M., & Szulecka, K. (1994). The Capgras and Cotard delusions. *Psychopathology*, *27*(3-5), 226-231.

Zeigler, H. P. (1964). Displacement activity and motivational theory: A case study in the history of ethology. *Psychological Bulletin*, *61*(5), 362.