What Empathy Can(not) Do. An Inquiry into the Epistemic Possibilities and Limits of Empathic Imagination.

Marco Inchingolo

(2021)

Examination Committee/Commissione di esame

Paul Bloom, Yale University

Jérôme Dokic, École des Hautes Études and Sciences Sociales

Cristina Meini, Università del Piemonte Orientale

Elisabeth Pacherie, École Normale Supérierure

Laurie Paul, Yale University

The copyright of this Dissertation rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

End User Agreement

This work is licensed under a Creative Commons Attribution-Non-Commercial-No-Derivatives 4.0 International License: https://creativecommons.org/licenses/by-nd/4.0/legalcode
You are free to share, to copy, distribute and transmit the work under the following conditions:

- *Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).*
- *Non-Commercial: You may not use this work for commercial purposes.*
- *No Derivative Works - You may not alter, transform, or build upon this work, without proper citation and acknowledgement of the source.*

In case the dissertation would have found to infringe the polity of plagiarism it will be immediately expunged from the site of FINO Doctoral Consortium

UNIVERSITÀ DEGLI STUDI DI GENOVA
UNIVERSITÀ DEGLI STUDI DI TORINO
UPO
UNIVERSITÀ DEGLI STUDI DI PAVIA

Acknowledgements

The long story of this work is entrenched with some of the most important chapters of my life. In this thesis, I distilled the knowledge I acquired throughout these past three and a half years. However, the final product is simply a pale reflection of the journey I made to produce it.

For this reason, I want at least to thank some of the fellow-travelers I met throughout this journey.

Grazie ad Andrea, la mia anima gemella e Andrea che ricopre lo stesso ruolo, ma a latitudini differenti. (A voi decidere quale amichetto ho menzionato prima)

Grazie alla mia famiglia allargata: Vito, “sire” fra tutti gli amichetti, Gimmy, il mio contrappunto spirituale, Francesca, che mi insegna il lato pragmatico della vita, Paola, che ormai è la mia seconda sorella, Silvia e Giacomo (Irene), che hanno visto muovere i primi passi di questo lavoro nel nostro nido torinese. Gege, Gianvito, Roberto, se scrivessi una frase divertente e intelligente per ognuno di voi non consegnerei mai questo lavoro. Vi voglio bene, sono fortunato ad avervi come amici.

Grazie a Marianna, il raggio di sole che ha illuminato gli ultimi mesi di questo lavoro.

Thanks to my Parisian family: Slawa, my long-lost brother and my all-time favorite German (and my all-time favorite Russian), to Pablo, for being the personification of kindness, to Antoine, for being such a lovable offspring of evolution, to Céline, a delightful friend and an irreplaceable intellectual partner.

Grazie alla mia famiglia “ristretta”, mio padre e mia madre. Senza il vostro supporto non avrei mai concluso niente. A mia sorella, si litiga davvero solo quando ci si ama davvero.

Thank you Joséphine. You have accompanied me throughout most of this journey. I cannot imagine doing any of it without you.

Thank you to my supervisor, Jérôme Dokic, who warmly welcomed me at the IJN and wisely guided me throughout these years.

Thanks to Laurie Paul, my fantastic guide at Yale. She managed to make me feel part of their community even on a screen, at hundreds and hundreds of miles away.

For those I forgot to mention, it’s only because I’m in a rush, not because I’m ungrateful. I’m sure you’ll understand. Thank you!

Table of Contents

ACKNOWLEDGMENTS 2

ABSTRACT 6

INTRODUCTION

BETWEEEN TWO EXTREMES 8

CHAPTER I. EMPATHY WHAT IS IT?

1. EMPATHY: SOME PRELIMINARY CONCERNS 12
2. FAMILIARIZING WITH THE DIVERSITY OF EMPATHY 13
3. DIFFERENTLY RELATED BUT NEVERTHELESS DISTINCT PHENOMENA 17
4. TOWARDS SOME KIND OF UNITY 18
   4.1. Towards Some Kind of Consensus 19
5. TOWARDS AN ACCOUNT OF EMPATHY 26
6. EMPATHY AND KNOWING WHAT IT'S LIKE: INTRODUCING PHENOMENAL INSIGHT 29
   6.1. Phenomenal insight: Two Preliminary Clarifications 30
   6.2. Imagining another's perspective via belief-like imagination 33
   6.3. Phenomenal Insight and Imagination: Experiential Embedding 36
   6.4. A Brief Clarification 39
7. DO WE REALLY NEED THE FIRST-PERSON PERSPECTIVE? 40
   7.1. Cappelen & Dever's Critique 41
   7.2. Why Cappelen and Dever's critique misses its goal 43
   7.3. Taking Stock: Perspectives Matter 50
8. A FEW MORE WORDS ON THE FIRST-PERSON PERSPECTIVE 50

ONWARDS 53

CHAPTER II. WHAT EMPATHY CAN DO: UNDERSTANDING REASONS (AND BEYOND)

1. INTRODUCTION 54
2. REASON = BELIEF + DESIRE? 56
3. RE-ENACTIVE EMPATHY AND UNDERSTANDING REASONS 58
4. ARATIONAL ACTIONS 60
   4.2. Goldie's Humean tenacity 62
   4.3. Stueber's Humeanism 63
5. ACCOUNTING FOR THE EXPERIENTIAL DIMENSION OF EMOTIONS 64
   5.1. Affective phenomenology and motivation 65

5.2. Back to phenomenal insight: Arational emotional actions 67
5.3. Back to phenomenal insight: Rational emotional actions 68
6.1. What about unconscious emotions? 70
6.2. Rationalization as a (useful) fiction? 75

ONWARDS 80

APPENDIX. WHAT IS IT (LIKE) TO IMAGINE AN EMOTION?

1. INTRODUCTION 82

2. THREE WAYS TO CONCEIVE OF THE RELATIONSHIP BETWEEN IMAGINATION AND EMOTION 83

3. SOME PRELIMINARY CONCERNS 87

4. SUBSTANTIATING EMOTION-LIKE IMAGINATION 90

5. FROM EMOTION REGULATION TO EXPERIENTIAL IMAGINATION 97

ONWARDS: BACK TO EMPATHY 100

CHAPTER III. WHAT EMPATHY CANNOT DO: THE BOUNDED SELF

1. INTRODUCTION: WHAT PESSOA GOT WRONG ABOUT THE IMAGINATION 102

2. INDEXICALITY AND THE “THIN” PSYCHOLOGICAL PROFILE OF THE SELF 104

3. THE SELF IN THE CONTEXT OF IMAGINATION 106

4. A “THICKER” VERSION OF THE SELF 107

4.1. More on the influence of our actual self on our imagined self 110
4.2. The actual self in empathy: the case of imaginative resistance 118
4.3. Some empirical support 123

ONWARDS 128

CHAPTER IV. WHAT EMPATHY CANNOT DO: TRANSFORMATIVE EXPERIENCES

1. INTRODUCTION 129

2. EPISTEMICALLY AND PERSONALLY TRANSFORMATIVE EXPERIENCES: INTRAPERSONAL AND INTERPERSONAL 131

3. Epistemic transformation 134

4. PERSONAL TRANSFORMATION AND GOLDIE’S ANTI-EMPATHY 140

4.1. Dispositions as the background of first-personal agency 141
4.2. Goldie’s critique and (personally) transformative experiences 145
4.3. Resisting Goldie’s critique: Imagining under constraints and the use of propositional knowledge for empathy 149
4.4. Explanation-aided understanding 154
4.5. Are Goldie’s and Paul’s concerns solved? 157
4.6. Taking stock 163

ONWARDS 164

CHAPTER V. WHAT EMPATHY CAN DO: LEARNING THROUGH EMPATHY, CHANGING THROUGH EMPATHY

1. INTRODUCTION 166

2. MOVING BEYOND THE BLACK-AND-WHITE MODEL OF EPISTEMICALLY TRANSFORMATIVE EXPERIENCES 167

3. A CLOSER LOOK TO THE PARENTING EXAMPLE: FROM AN EXPERIENTIAL "GESTALT" TO AN EXPERIENTIAL "COMPOUND" 169

4. TROUBLES WITH PERSPECTIVE-TAKING 171

4.1. Egocentric bias 172
4.2. Affective forecasting 174
4.3. Getting perspectives rather than taking perspectives? 177

5. GAINING INSIGHTS INTO NEW EXPERIENCES 181

5.1. Phenomenal imagination, phenomenal properties, phenomenal concepts 183
5.2. Noticing 184
5.3. Comparisons and involuntary memory 185
5.4. From literature to empathy 187

6. MORE ON EMPATHY AND MEMORY 188

7. PERSONAL OUTCOME: CHANGING THROUGH EMPATHY 192

7.1. THE EMPATHIC FRAMING OF INFORMATION 193

7.2. Brief clarificatory interlude 197
7.3. Empathy and other ways in which the world can be 200

CONCLUSIONS

END OF THE JOURNEY (LOOKING AHEAD) 203

BIBLIOGRAPHY 206

Abstract

My dissertation is an exploration of the epistemology of empathy, namely of some of its major epistemic possibilities and limits. Throughout the present work, I am going to argue that empathy is a cognitive process through which we imagine another person's first-person perspective in a given situation. Empathy aims at providing empathizers with an insight into "what it feels like" to be in somebody else's shoes.

The dissertation is so structured.

In *Chapter I*, I begin in an admittedly unexciting way, hence by clarifying the terms I will deploy. However, as we will see, this effort of clarity is useful – if not necessary – to avoid undue confusion whenever we decide to engage with a topic like empathy. I will engage with the large literature devoted to the very definition of empathy. I discuss the many possible accounts of it one can find in both the empirical and philosophical literature. From this debate, I extrapolate a pattern of basic features that are shared by most accounts of empathy. In light of this, I then introduce the definition of empathy I will be dealing with throughout the present work: empathy as a process of imaginative perspective-taking aiming at *phenomenal insight*, i.e. at first-personally representing what *it feels* like to be in somebody else's shoes.

In *Chapter II*, I engage with Karsten Stueber's account of empathy as providing us with an understanding of other people's reasons for action, where reasons are to be taken as mainly constituted by a belief-desire couple. I criticize the belief-desire model and argue for the inclusion of emotions as full-rights reason-giving states. In order to account for the reason-giving dimension of emotional states, I argue that we need to ascribe to empathy the capability of providing us with an insight into the very experienced dimension of mental states. Indeed, I will show how emotions' main source of motivation can be found in the way in which they are phenomenally experienced by subjects. I then integrate Stueber's account of empathy with my notion of phenomenal insight to include emotions in what can be empathically re-enacted in order to understand how they can motivate us to act.

Attached to this chapter I put an *Appendix* in which I explore more in detail how it is possible for us to first-personally imagine emotional states. I will offer a typology of several ways in which it is possible to imagine emotional states and individuate a particular type of imaginative process that seems to be particularly apt for our purpose, namely emotion-like imagination. I then sketch the conceptual profile of emotion-like imagination and provide an empirically informed analysis of the psychological mechanisms underlying it by resorting to the research on emotion regulation.

In *Chapter III*, I argue that what can be empathically imagined is *bound* to the kinds of individuals we are, i.e. our preferences, values, dispositions, etc. When trying to imagine other people's perspectives, traces of our "selves" can be found in what we imagine. That is likely to considerably constrain the range of perspectives we can empathically imagine. In order to better clarify my point, I will address the issue of the so-called "imaginative resistance", which I take to be best explained in terms of an intrusion of our "self" in the imaginative (and, therefore, empathetic) process. I finally back my claims by referring to some relevant empirical work in social psychology and neuroscience.

In *Chapter IV*, I explore further how empathy is linked to the kind of people we are by engaging with the contemporary debate on transformative experiences (TEs). The implication TEs on our capacity to perspective-take has been rarely tackled as such as the debate has tended to focus on the self-disrupting dimension of TEs and its implication for prospection and decision-making. Here, I will mainly focus on issues pertaining to perspective-taking (viz. empathy), rather than decision-making, by asking questions such as: how can I imagine experiences I did not have? And, how can I imagine being a different person? TEs are particularly interesting for the purpose of the present work for they seem to deepen the point I raised in the previous chapter about the bounded self. The bundle of issues related to TEs could, indeed, be interpreted as highlighting the limitations of our imaginative (or, more broadly, representational) capabilities due to the kind of self we are. The stock of experiences we had constrains our capability to conjure up the relevant imaginings about kinds of experiences we did not personally undergo. At the same time, the kind of individuals we are impedes us to fully appreciate a different self's perspective – with other dispositions, another system of value and preferences, etc. In other words: who we are shapes and constrains our empathic processes.

In *Chapter V*, I am going to see how the challenges to empathy explored in the previous chapter can be counterbalanced. Rather than trying to show how Paul's challenges to empathy can be completely overcome, I will try to show how they can be *faced* by empathizers. This will allow me to show some surprising features of empathy that are seldom discussed in contemporary literature, namely the possibility to *learn* via empathy and the possibility to *change* via empathy. On the one hand, by stretching our imagination, empathy can provide us with phenomenal insight into experiences we did not actually undergo. Empathy can, therefore, deliver relevant knowledge about perspectives that are radically different from our own. In other words, we can learn through empathy. On the other hand, I am going to argue that empathy, by exposing ourselves to new perspectives, can *change* us. In other words, by bringing ourselves to appreciate different worldviews, empathy can re-shape the kind of people we are.

Introduction
Betweeen Two Extremes

In *The Lives of Animals*, Elizabeth Costello, a fictional character created by the mind of J. M. Coetzee, while speaking at a conference held at a prestigious institution in New England, argued the following:

> [T]here is no limit to the extent to which we can think ourselves into the being of another. There are no bounds to the sympathetic imagination. […] I can think my way into the existence of a bat or a chimpanzee or an oyster, any being with whom I share the substrate of life. (Coetzee 1999/2016: 35)

In a passage of *In Search of Lost Time,* French writer Marcel Proust wrote:

> A real person, profoundly as we may sympathise with him […], remains opaque, presents a dead weight which our sensibilities have not the strength to lift. If some misfortune comes to him, it is only in one small section of the complete idea we have of him that we are capable of feeling any emotion. (Proust 1913/1992: 116-117)

Mine is obviously not a work of literary scholarship, but we can learn a lot about the topic discussed in the following pages through the words of these two great writers. The topic I am alluding to is imagination, specifically that kind of imagination that allows us to explore perspectives that are different from the one we are currently occupying. I will call this process of imaginative transposition, empathy.

As we will see, the very notion of empathy is intrinsically volatile and hotly debated. I will argue that empathy is that particular imaginative process through which we "think our way into the being of another". In other words, empathy is what allows us to first-personally imagine ourselves in the shoes of another, viz. to imaginatively take up the cognitively centered perspective of another subject.

Back to where we started. I take Coetzee's and Proust's quotes to be diagnostic of a certain tension that inhabits the very concept of empathy. This tension is what will animate the pages to come. It concerns empathy's epistemology, namely the kind of knowledge and understanding that it can deliver about other people's minds. More precisely, the tension that I take to be essential to empathy concerns its epistemic possibilities and limits.

On the one hand, empathy seems to be an invaluably powerful tool to gain insight into the lived experience of someone else. As Coetzee writes, there are *no limits* to what he calls (following Thomas Nagel) the "sympathetic imagination" (and what I decided to call empathy). We can surf across a seemingly endless variety of perspectives and think our way into all beings with whom we share the "substrate of life". Empathy, as portrayed by Coetzee (*alias* Elizabeth Costello), can bridge any kind

of divide between conscious beings. Not only can it bridge the gap that separates a person from another but it can also erase those that separate our species from others – bats and oysters included. It is the ultimate medium to establish a boundless intersubjective community of souls.

On the other hand, there seems to be limits to what we can achieve via empathy. The understanding that our imagination can bestow about other minds is never going to be exhaustive. With the words of Proust, other people *remain opaque*. No matter how tenacious our attempts may be, there will always be aspects that our empathic grasp will miss. As we will see throughout the present work, empathy's limits can run even deeper than that. Empathy, indeed, can not only miss some crucial aspects of other people's perspective but can go astray in multiple ways. In other words, empathy's understanding can be not only be *incomplete*, but also *biased* and intrinsically *defective*.

In the present work, I try to place myself in between these two extremes. Sitting with Proust on my right side and Coetzee on my left side to explore the space in between. I will survey what empathy can and cannot do, i.e. what we can and in fact do understand via empathy and what we cannot and in fact fail to understand via empathy.

In my intellectual quest, I will abandon the poetic tones of literature in favour of those of philosophy and cognitive science. I will resort to conceptual analysis and back, whenever deemed appropriate, my considerations with empirical results.

It would be hopelessly naïve to believe that my analysis will be definitive and exhaustive, and I aim at neither of those goals. What I will do is rather addressing some issues that I deem particularly crucial for empathy as they offer a privileged angle from which to highlight its conceptual profile and the epistemic outcomes it entails.

The present dissertation is so structured.

In *Chapter I*, I start in an admittedly unexciting way, hence by clarifying the terms I will deploy. However, as we will see, this effort of clarity is useful – if not necessary – to avoid undue confusion whenever we decide to engage with a topic like empathy. I will engage with the large literature devoted to the very definition of empathy. I discuss the many possible accounts of empathy one can find in both the empirical and philosophical literature. From this debate, I extrapolate a pattern of basic features that are shared by most accounts of empathy. In light of this, I then introduce the definition of empathy I will be dealing with throughout the present work: empathy as a process of imaginative perspective-taking aiming at *phenomenal insight*, i.e. at first-personally representing what *it feels* like to be in somebody else's shoes.

In *Chapter II*, I engage with Karsten Stueber's account of empathy as providing us with an understanding of other people's reasons for action, where reasons are to be taken as mainly constituted by a belief-desire couple. I criticize the belief-desire model and argue for the inclusion of emotions as full-rights reason-giving states. In order to account for the reason-giving dimension of emotional states, I argue that we need to ascribe to empathy the capability of providing us with an insight into the very experienced dimension of mental states. Indeed, I will show how emotions' main source of motivation can be found in the way in which they are phenomenally experienced by subjects. I then integrate Stueber's account of empathy with my notion of phenomenal insight to include emotions in what can be empathically re-enacted in order to understand how they can motivate us to act.

Attached to this chapter I put an *Appendix* in which I explore more in detail how it is possible for us to first-personally imagine emotional states. I will offer a typology of several ways in which it is possible to imagine emotional states and individuate a particular type of imaginative process that seems to be particularly apt for our purpose, namely emotion-like imagination. I then sketch the conceptual profile of emotion-like imagination and provide an empirically informed analysis of the psychological mechanisms underlying it by resorting to the research on emotion regulation.

In *Chapter III*, I argue that what can be empathically imagined is *bound* to the kinds of individuals we are, i.e. our preferences, values, dispositions, etc. When trying to imagine other people's perspectives, traces of our "selves" can be found in what we imagine. That is likely to considerably constrain the range of perspectives we can empathically imagine. In order to better clarify my point, I will address the issue of the so-called "imaginative resistance", which I take to be best explained in terms of an intrusion of our "self" in the imaginative (and, therefore, empathetic) process. I finally back my claims by referring to some relevant empirical work in social psychology and neuroscience.

In *Chapter IV*, I explore further how empathy is linked to the kind of people we are by engaging with the contemporary debate on transformative experiences (TEs). The implications of TEs on our capacity to perspective-take have been rarely tackled as such as the debate has tended to focus on the self-disrupting dimension of TEs and its implication for prospection and decision-making. Here, I will mainly focus on issues pertaining to perspective-taking, rather than decision-making, by asking questions such as: how can I imagine experiences we did not have? And, how can I imagine being a different person, i.e. having preferences, values, dispositions I do not actually have? These are *intra-personal* issues as much as they are *inter-personal*: they have implications on both how we think of ourselves in the future and at the same time impact how we *understand* alien perspectives and individuals who are significantly different from us.

TEs are particularly interesting for the purpose of the present work for they deepen the point I raised in the previous chapter about the bounded self. The bundle of issues related to TEs could, indeed, be interpreted as highlighting the limits of our imaginative (or, more broadly, representational) capabilities due to the kind of selves we are. The stock of experiences we had constrains our capability to conjure up the relevant imaginings about kinds of experiences we did not personally undergo. At the same time, the kind of individuals we are – i.e. our preferences, dispositions, values we have – impedes us to fully appreciate a different self's perspective – with other dispositions, another system of value and preferences. In other words: who we are shapes and constrains our empathic processes.

In *Chapter V*, I am going to see how the challenges to empathy explored in the previous chapter can be at least counterbalanced. The arguments I present in this chapter and those offered in the previous one are not mutually exclusive. Indeed, rather than trying to show how Paul's challenges to empathy can be completely overcome, I will try to show how they can be *faced* by empathizers. This will allow me to show some surprising features of empathy that are seldom discussed in contemporary literature, namely the possibility to *learn* via empathy and the possibility to *change* via empathy. On the one hand, by stretching our imagination, empathy can provide us with phenomenal insight into experiences we did not actually undergo. Empathy can, therefore, deliver relevant knowledge about perspectives that are radically different from our own. In other words, we can learn through empathy – and not only through experience. On the other hand, I am going to argue that empathy, by exposing ourselves to new perspectives, entails the possibility to *change* us. In other words, by bringing ourselves to contemplate different worldviews, empathy can re-shape the kind of people we are.

Chapter I

Empathy: What is it?

1. Empathy: Some Preliminary Concerns

One of the aspects that are likely to impress someone who is approaching the topic of empathy for the first time is the rich disciplinary panorama within which this phenomenon is investigated. Empathy prominently figures in large-audience books and scientific publications in philosophy, psychology, neuroscience, anthropology, political science, aesthetics, law studies, and literary studies, just to mention a few. The phenomenon is regularly examined through different methodological lenses and diverse conceptual frameworks.

In light of this variety of approaches, one has to expect an equally rich variety of interpretations. Depending on the particular disciplinary "take" scholars from various fields have on the topic, we see different features of the phenomenon being highlighted and some others, which are further away from the epistemic scope of the discipline, being put on the background or utterly ignored. As predictable, while on the one hand this disciplinary diversity surely improves our overall understanding of the phenomenon, on the other hand, it also generates some disagreement about the very nature of empathy. Dispersed among and contended by different disciplines, empathy's conceptual identity is constantly in jeopardy.

Unfortunately, the situation is even worse than that. Disagreement is indeed not only to be found *across* different disciplines but also – and massively – *within* the same discipline. Indeed, the definitional boundaries of empathy are hotly debated among scholars in the same discipline. And even when definitional concerns are not explicitly addressed, scholars within the same field deploy the term "empathy" quite liberally. As a result, their research-outputs, while falling under the same terminological umbrella, actually concern different phenomena. This is likely to exacerbate confusion and increase the risk that conflicting views on empathy merely amount to verbal disputes among people that are actually talking about different things.

On top of this, one has to be aware of the pervasive use of empathy in the context of our everyday life. Extremely common are indeed sentences such as "he is a really empathic person", or "she totally lacks empathy", or "can't you be a bit more empathetic?", etc. This implies that virtually everyone has at least a rough idea of what empathy is or what empathy is supposed to do or what it means to be empathic.

For all these reasons, scholars interested in empathy and in assessing its role in the human practice of mutual understanding, face a serious complication as soon as they approach the topic. They have indeed to navigate a crowded area of research where diverging accounts, methodological approaches, conceptual tools, and folk intuitions are one of the distinctive marks of the phenomenon investigated.

In light of that, some of the preliminary questions we need to take care of before delving into the main theoretical aim of the present dissertation are the following ones: what is empathy? Is it a homogeneous phenomenon? How is it possible to navigate among so many diverging approaches? Will it ever be possible to reach some form of consensus about the definition of empathy?

I will not provide definitive answers to these questions. I will rather address the legitimate epistemic worries they express. As we will see, I do not aim to solve any rigorous definitional disputes about empathy. What I am interested in is clarifying the conceptual landscape I will be dealing with in order to avoid terminological and conceptual confusions.

In what follows, I will make clear the *kind* of empathy that will be scrutinized in this work. At the same time, though, the kind of empathy I will be referring to is not the result of a mere stipulation. I will not simply group together some random phenomena I find interesting and then stick the "empathy" label on them. Rather, I will motivate my notion of empathy by integrating it within a constellation of phenomena in social cognition that many scholars in the field would not hesitate to call empathetic. This will lead us to individuate some underlying unity in the conceptual galaxy of the various things we usually call empathy.

2. Familiarizing with the diversity of empathy

There have been several attempts to map the different uses of empathy in numerous scientific publications. These attempts are either supposed to provide precious orientation tips (e.g. Batson 2009), intuitive evidence for the need for a more rigorous definition of empathy (e.g. Coplan 2011). In our case, they will help us familiarize ourselves with the diversity of empathic phenomena.

A very simple example might turn out useful to set up the stage. Let us imagine a real-life case which is more than likely most of us experienced in some way. A friend of yours comes towards you, she is weeping. She just broke up with her partner after a long and intense relationship. Tears stream down her face while she tells you the unfolding of their last conversation and how overwhelmed by frustration and anxiety she feels whenever she attempts to picture her life without him from now on.

You might undergo a variety of experiences and engage in a wealth of mental processes while facing a situation like this. Some of these processes could be easily identified as empathetic.

Let us try to list very briefly at least some of the most prominent.

N.B. The following list is not meant to be exhaustive but simply to endow us with some preliminary orientation cues. Also, I will not address the specificity of every process enlisted because each of it would require a monographic work of its own.

i) Emotional Contagion

While listening to your friend pouring out her heart to you, you might come to “catch” her emotions as they get expressed by her gestures, behavior, intonation, facial expression, etc. By simply being close to her, you might become receptive to her emotional state and come to experience it yourself. In one word, your friend’s emotion becomes “contagious”, i.e. capable of spreading around the people who are close to her and paying enough attention to her condition.

This phenomenon is usually called “emotional contagion” and its occurrence is supported by extensive neuroscientific evidence. From a strictly neurological perspective, emotional contagion occurs when remarkably overlapping brain areas are activated when subjects experience the emotions “for themselves” or when they perceive an emotion displayed by someone else[1].

Interestingly enough, given its quasi-automatic nature of this process, emotional contagion can happen without any substantial form of interpersonal understanding taking place. Indeed, I might come to experience sadness (or some other kind of negatively valenced emotional state) without being aware that it comes from the other person (Goldie 2000). Moreover, the emotion that we contagiously caught from the other person can likely become our own and make us blind about the other person’s emotion by drowning us into the downward spiral of our own affective life (Prinz 2011). To stick with the example we are focusing on, the emotion of sadness we catch from the other person is likely to direct our attention towards our personal reasons for being sad. For instance, we might be reminded of our recently finished relationship and caught up in thinking about it. As soon as we do that, we might cease to pay attention to what our friend is saying.

[1] As neuroscientific reviews of the phenomenon see (Hatfield, Rapson, & Le, 2009); and (Bernhardt & Singer, 2012)

ii) Affective Sharing

Affective sharing, also known in the literature as "affective empathy", is in many ways related to emotional contagion. Indeed, it also consists in coming to catch the emotional state of the other person. At the same time, affective sharing comes with a *plus* if compared to mere emotional contagion. It is taken to bear with it a more substantial form of interpersonal understanding. Indeed, the main crucial difference from instances of mere emotional contagion lies in the awareness of the self-other differentiation that accompanies the process. Through affective sharing, I am supposed to experience myself the emotion of the other, while still being aware that I am *vicariously* experiencing it. In this case, we could say that my emotional state is outwardly oriented and constitutes a form of emotional receptivity that attunes us with the other person's emotional states without, so to speak, "switching the roles" between self and other.

iii) Knowing what she is feeling/thinking

Being capable of accurately inferring what the other person is feeling or thinking is usually called in the empirical literature "cognitive empathy" or "empathic accuracy" (Ta and Ickes 2017).
This phenomenon sticks out for its conceptual broadness. Indeed, our capability to accurately assessing the thoughts and feelings of other individuals is something that can be accomplished through a variety of cognitive processes and inferential strategies. Therefore, if we spell out empathy in these terms, we decide to set aside questions pertaining what kind of cognitive processes empathy is supposed to instantiate. Moreover, this notion would simply conflate the notion of empathy with that of mindreading, i.e. our capacity to accurately track other people's mental states.

At the same time, though, this notion is widely used in the psychological literature and has been popularized by a series of very influential studies run by William Ickes and colleagues on empathic accuracy (Ickes 2003). One of the most known ways in which Ickes and colleagues test empathic accuracy consists in filming a social interaction between two participants. The video is then shown to them and each participant is asked to write down what they were feeling or thinking at specific moments of the interaction. Later on, they have to watch the video again and try to infer what *the other persons* were thinking or feeling at the specific intervals where they reported they were thinking or feeling something in particular. With the words of William Ickes Vivian Ta:

> Empathic accuracy, as operationally defined, is a global measure of how accurately a perceiver is able to infer the specific content of all of the target person's thoughts and feelings. The accuracy of the perceiver's empathic inferences is based on trained raters' judgments of how similar each of the inferred thoughts and feelings were to the actual thoughts and feel- ings that the target person reported […].The global (overall) measure of empathic

> accuracy is scaled as a percentage measure, with 0 indicating no accuracy and 100 indicating perfect accuracy. (More precisely, 0 indicates that the perceiver was awarded no "accuracy points" for any of his or her empathic inferences, whereas 100 indicates that the perceiver was awarded all of the available "accuracy points" by the trained raters.) (Ta & Ickes 2017: 354)

Coming back to our guiding example, while interacting with my friend, I might come to infer a variety of mental states she is likely to experience during our conversation. If, say, at some point she is feeling "angry" and I believe she is feeling "angry", then my "accuracy points" would be 100, if I believe she is "happy", my accuracy points would be 0.

iv) Perspective-Taking

We are still with our friend. She is still telling us about her failed relationship. One thing that we are very likely to do in such circumstances is trying to put ourselves in her shoes, i.e. we can try to imaginatively adopt her perspective to figure out how things might look from there. Perspective-taking processes are also known, especially in the philosophical literature, as "simulation" (Barlassina and Gordon 2017; Currie and Ravenscroft 2011; Goldman 2006; Heal 2003; Stueber 2006) and is sometimes called cognitive or imaginative empathy (Spaulding 2017).

There are several ways in which I can imaginatively work out somebody else's perspective. A distinction that is often presented in the psychological literature and that has produced a fair amount of evidence is the distinction between "imagine-self" and "imagine-other" (e.g. Maibom 2020).

In the imagine-self case, what I do is merely imagining what *I* would do if I were happened to be in my friend's shoes. That is, given the kind of individual I am, the kind of beliefs, preferences, aims, personality traits I have, I try to figure out how I would react if I were to face the end of an important relationship. In the imagine-other scenario, I would rather try to imagine what *my friend* would do. That is, based on my knowledge about the kind of person she is, about her beliefs, preferences, aims and the specificity of her character traits, I try to figure out how she might feel in a specific situation. The wider the mismatch in the interpersonal equation, i.e. the greater the differences between me and my friend, the more likely it is that the two imaginative processes produce diverging outcomes.

As we will try to show later on in the present work (*Chapter III*), even if the distinction is intuitively appealing – few would deny it – and solidly even if this difference is backed by evidence in the empirical literature**,** it remains partly unsatisfying. The main reason why it is so is that, in many contexts, it might turn out to be too artificial. If we indeed look at our everyday practice of perspective-taking, there would be few instances of purely imagine-self or purely imagine-other, but always a mixture of the two.

These are all matters that will be discussed more at length later on in the present work, we will try to assess the role of the self in processes of imaginative empathy. For now, though, it suffices to be aware of this bundle of issues.

v) Concern, Sympathy, Compassion

Finally, when seeing our beloved friend in such a miserable and painful condition, we might feel what is called "empathic concern" for her. Differently from emotional contagion and affective sharing, we respond with an emotion that is not necessarily *isomorphic* to what she is experiencing. For instance, supposing that she is feeling anger or resentment, we do not vicariously experience a similar type of emotion, but rather experience an emotion that reflects an overall caring attitude towards her (Maibom 2020). By coming to appreciate her suffering we "bring home […] [her] case to our own bosoms"(Smith 1759/2009: 269), that is we come to feel concern, sympathy, or compassion *for* her.

These kinds of empathic phenomena are undoubtedly moral forces. Indeed, by allowing us to emotionally "tune-up" to someone else's suffering, we are – at the very least – disposed to take action to improve their well-being. There can be differences in how one precisely characterizes notions such as empathic concern, compassion, or sympathy, but surely they seem to share a fundamental characteristic: they can all be conceived as affective devices for moral attunement that are likely to dispose us to act in ways that are functional to the other person's overall well-being.

3. Differently related but nevertheless distinct phenomena

In the previous section, we were able to individuate certain broad categories of empathic phenomena in virtue of some coarse-grained features that allowed a rather straightforward differentiation among them.

If we look more carefully, though, we can see how, in real-life social interactions, these processes are intimately related to each other. It was not a coincidence, indeed, that we were capable of intuitively illustrating some of the distinctive features of each phenomenon by means of a singles example (our friend's telling us about her recent breakup).

We can indeed see how a process of mere emotional contagion could easily evolve in a case of affective sharing as soon as the subject vicariously experiencing the emotion becomes also aware of the source of her actual emotional state. Or we can see how, by imagining being in somebody else's shoes, we can come to experience the very same emotion she is directly experiencing. Or by means

of perspective-taking or affective sharing, we can come to experience concern and sympathy for the other person and adopt a caring attitude towards her.

To put it roughly, we can see how every phenomenon can easily slip into the other so that those previously enlisted empathic phenomena, while being conceptually distinct, are intertwined and intimately related to each other.

However, this awareness should not make us blind to the philosophical importance of being mindful of their crucial differences. Indeed, even if empathic phenomena usually are likely to occur together in our everyday experience, they can (and in fact *do*) also come apart from each other. Being aware of this possible divergence is of crucial philosophical and scientific importance.

For instance, if we are trying to assess the role of empathy in motivating prosocial behaviours, we need to be able to distinguish between emotional contagion and perspective-taking. Being oblivious of this difference and conflating the two notions in an indistinct empathic "primordial soup", could result in the philosophically sloppy conclusion that perspective-taking entails emotional contagion and in an empirically awfully confounded correlation between empathy and prosocial attitudes: indeed we would not be able to control for what is actually doing the prosocial job, perspective-taking or affective sharing.

Therefore, conceptual distinctions, if kept to a reasonable level, are far from being mere obsessive hair-splitting manias but serve the purpose of rigorous philosophical speculations and sound experimental designs.

4. Towards Some Kind of Unity

Our brief venture in the multifarious landscape of empathy hopefully succeeded in making us aware of some coarse-grained distinctions among sets of phenomena that scholars have identified as empathic. As already announced though, one needs to be aware that phenomena enlisted in the previous section not only do not capture *everything* that scholars have called empathy but also omit some fine-grained differences that can be individuated *within* the same category.

I believe it would be philosophically futile (and, frankly, somewhat boring) to chase after all the tiny differences between various empathic phenomena. Also, there are numerous valuable publications

out there that brilliantly succeed in giving comprehensive overviews on empathy, so that replicating another one in the present work would not add much to the contemporary debate.[2]

Therefore, if the previous section tried to isolate certain conceptual macro-areas about empathy in the attempt to get a grip on the bigger picture, in the present section I will try to move some steps towards the conceptualization of the notion of empathy that will be deployed throughout this work.
As anticipated, my argumentative strategy will not be purely speculative, i.e. I will not merely present a notion of empathy that can be more or less straightforwardly ascribed to one of the categories enlisted above.

I will rather try to work towards the individuation of some kind of unity among various empathic incarnations and then zoom-in to circumscribe the notion of empathy that will concern us throughout the present work.

Indeed, as rightfully noted by (De Vignemont, Jacob 2017: 511) "a certain amount of stipulation is unavoidable in the way one uses quasi-technical terms such as empathy". For that reason, the arguments to follow have no strong normative aims, i.e. I do not intend to show how scholars in the field *should* adopt a conception of empathy that is in line with mine.

However, I also believe that motivating one's choice for a certain kind of empathic phenomenon over another is a philosophical plus since it allows is to rigorously individuate the epistemic scope of the phenomenon inquired and illustrate empathy's specific contribution to the realm of human social cognition.

Therefore, I will point out at some good reasons we have to take empathy to have the specific features I will highlight in what follows. This strategy allows me to do two things: i) motivate my choice for this specific account of empathy in a non-purely stipulative way and at the same time ii) highlight and illustrate certain distinctive features of my notion of empathy.

4.1. Towards Some Kind of Consensus

In a recent publication, philosopher John Michael (Michael 2014) reviewed some of the most influential accounts of empathy in the attempt to individuate a cluster of shared features that would allow scholars in the field to gather around a space of minimal consensus about the conceptualization of empathy.

---

[2] For some recent excellent overviews on empathy see at least (Maibom 2020) and (Matravers 2017).

According to him one of the fundamental reasons of conflict is due to the fact that some researchers regard as necessary features of empathy what some others tend to interpret as mere characteristic features or as typical consequences of it.

From this consideration he draws the following conclusion: notwithstanding the large disagreement, different scholars with competing views about empathy do show some sort of consensus towards its nature, which is ultimately what allows for meaningful comparisons among diverging views.

From Michael's analysis we learn that some of the most influential scholars in the field tend to agree on some basic principles that seem to generally regulate empathy's cognitive mechanisms and epistemic outcomes. Divergences emerge in the way scholars in the field spell out how empathy incorporates these fundamental principles.

Following Michael, we can extrapolate that empathy usually revolves around the following four elements: (a) interpersonal understanding, (b) engagement, (c) similarity, (d) defeasible caring. Let me address them in turn.

a) *Interpersonal Understanding*

One of philosophy's biggest questions is known as the "problem of other minds" (Avramides 2019). So formulated, it seems like there is only one kind of problem regarding other minds, but actually there are at least two versions of it: the skeptical doubt about other minds and the problem of mindreading.

The first version of the problem has an illustrious tradition that dates back at least to Descartes and his concern about the existence of the external world, hence of the other minds as well. Skepticism, brought to its extreme consequences, leads philosophers to this dramatic concern: given that I have direct access only to the contents of my own mind, how can I be so sure that other individuals have minds? Could it be that what I take to be people with minds, just like mine, are in fact disguising zombies with no soul?

The mindreading version of the problem of other minds is, so to speak, less solemn and more mundane, even if no less important than the first one. Instead of putting the existence of other minds into question, it asks: given that other people around me do have minds, how do I figure out what is going on inside them?

Empathy is an answer – one among many others, of course – to the second version of the problem. Very few scholars in the field would argue against a conception of empathy as a process that enables some kind of insight into the minds of other individuals.[3]

b) *Engagement*

The notion of engagement is a rather elusive one. Indeed, taken out of context, it might look as both extremely abstract and metaphorical. Therefore, it needs some philosophical labour to clarify its meaning.

To that end, it might be useful to recall the distinction operated by the German psychiatrist Karl Jaspers which has been recently recalled by John Campbell (Campbell 2015)**.** Referring to the kind of understanding a therapist can show towards her patient, Jaspers distinguished between a scientific explanation and an imaginative-empathic understanding.

With scientific explanations, the therapist refers to some previously acquired body of knowledge and competences (e.g. through psychiatry books) that can enable a *causal* reconstruction of the feelings and behaviors of the patient. In this case, the psychiatrist's understanding would amount to purely descriptive, law-like characterization of her patient's psychological state.

We can imagine, though, another way of achieving knowledge about the patient's mind that is in principle still available to the therapist, namely imaginative understanding – which we can take as simply another name for what we called in the previous section's list "perspective-taking". In the case of imaginative understanding, the therapist employs *her own* cognitive and emotional resources in order to disclose her patient's perspective. Here, there is no explicit reference to some law-like description regarding the functioning of the human mind and behavior, but rather the attempt to put

---

[3] Of course, this does not mean that *every account* of empathy will take it to instantiate some form of interpersonal understanding. As announced, in the present section, I am trying to narrow down the scope of empathic phenomena in order to achieve greater clarity and better introduce the account that will be deployed throughout the rest of the present work. This will inevitably *leave something out*: it will be impossible to keep everything that is called empathy out there within the narrower picture I am trying to sketch alongside Michael's insights. Taking a look back at the list in the previous section, surely the notion of emotional contagion would be left out. We saw indeed how poor of a guide to understanding others pure emotional contagion can be. We even saw how emotional contagion can come independently of any substantial kind of interpersonal understanding having to occur. The notions of care, concern, sympathy instead, are not something that centrally leads to interpersonal understanding but actually presupposes it. Anyway, as said, if we choose to somehow narrow down the scope of empathy, then it will be impossible to account for everything people call empathy.

oneself in the other person's shoes in order to achieve something that could be described as a first-personal understanding of the other's experiences.

We can easily see how these considerations can be generalized outside the mere domain of psychiatric therapy and extended to different cognitive strategies that do not exclusively reduce to the imaginative understanding mentioned by Jaspers.

Karsten Stueber (2006) individuated two general stances humans can take to make sense of other people's minds and behaviors: a *detached stance* and an *engaged stance*. To keep things simple, adopting a detached stance when mindreading others implies making reference to say a general set of principles regarding the usual functioning of human minds and behavior and draw informed inferences on the basis of these principles, given the evidence that is available to us at the moment we formulate our judgment. Going back to the example of the previous section: based on the evidence we have concerning the circumstances our friend finds herself in and some generalizations about how people tend to feel and react in these kinds of circumstances, we can make informed inferences about the goings-on of her mind. So, given that she just ended a very important relationship and since we know that people tend to be upset by failures of romantic relationships, we can infer with a certain degree of confidence that she must feel upset.

This strategy shares a similar logic to the scientific understanding mentioned by Jaspers. Moreover, it is largely in line with what is known in the relevant literature about social cognition as the "theory-theory". Theory-theory maintains that we make sense of other people's minds and behavior by basing our judgments on a set of naïve principles about the general functioning of human psychology.

We can easily see why this can be characterized as a *detached* approach to the problem of interpersonal understanding: it is a strategy that, in principle, even "alien minds with a completely different psychology could adopt for the purpose of explaining human behaviour" (Stueber 2006: 22), i.e. it is a strategy that could be deployed even if we are completely alien to (or *detached from*) the actual circumstances the other person finds herself into.

We can easily contrast this detached stance with an approach that *engages* us more directly in the process of understanding. We could indeed be first-personally *engaged* by actual experience of our target so that we can come to appreciate some fundamental features of it, so to speak, on our own skin.

This could be accomplished through Jaspers' imaginative understanding, where we attempt to directly engage our cognitive and affective resources to produce imaginative replicas of the other's

experiences. But it easy to conceive of other ways in which this more engaged stance could be successfully instantiated. If we look back at our list in the previous section, we can see that any form of affective sharing aligns to an engaged ideal of mutual understanding. Indeed, if we ourselves come to vicariously experience the same affective state of the other person, then it is clear that we are personally *engaging* with her goings-on, i.e. we are no more occupying an alien or external perspective to her mental life but are somehow caught up in it.

The only account presented in the previous list that would not fit with the "engagement" principle would probably be the one that equates it to mindreading. Indeed, if we simply take empathy to be any kind of accurate inference about the other person's thoughts and feelings, then both engaged and more detached approaches could be easily count as empathic, as far as they are accurate.

If we want to move within the epistemic coordinates that we are putting forward with the help of Michael's analysis, then we necessarily have to narrow down the notion of empathy to designate something more specific than any instance of inferential accuracy about the other person's mental states.

c) *Similarity*

This feature stems almost spontaneously from the previous one. As a general principle, the similarity condition claims that some sort of similarity between the target's and the empathizer's states needs to be in place for empathy to be successfully instantiated.

In the case of affective sharing this similarity strikes us as immediately evident: if the empathizer comes to experience a vicarious emotion as a response to the target's actual state and circumstances, then it is plainly true that some sort of interpersonal similarity is instantiated.

But some sort of similarity could be instantiated through multiple means. Coming back to what Jaspers and Campbell called "imaginative understanding", here as well some form of similarity needs to be in place. Indeed, in order to accurately represent the situation another person finds herself in, the empathizer needs to imaginatively reproduce at the very least a similar *intentional structure* of the target's experiences.

The similarity condition is not necessarily instantiated by the accounts of empathy that I grouped under at the point (iii) and (v). For what concern (iii), accurately inferring other minds' goings-on is something that can be achieved without any kind of similarity being in place. Accurate inference about other people's mental states is something that could be accomplished even by a mindreading technologic device. We can indeed imagine a processor that is capable of scanning humans' brain

activity and produce as outputs accurate labelling of their mental states. In this case, inferential accuracy would be achieved without any kind of similarity being instantiated between the processor and the target, whose mental life is incomparable to the workings of a processor.

The phenomena enlisted at point (v) also fall outside the scope of the principle of similarity. It is indeed not entailed by any of the notions of care, sympathy and concern that they be isomorphic to the emotional state of the other person. I can indeed feel sympathy for the other person's suffering even if there is very little similarity between the empathizer's and the target's states.

One concern one might have at this point regards what kind and what degree of similarity need to be in place in order for a state to be legitimately called empathetic: would the empathizer need to instantiate the exact same state of the target or will any similarly valenced affective state be enough? This is a very delicate matter that will crucially depend on the specificities of the account of empathy one is deploying. Therefore, it is not something that should concern us at this stage.

Now, as already stated, we simply need to come to appreciate some overarching principles that seem to underlie most accounts of empathy that we may find out there in the scientific work of numerous scholars.

d) *Defeasible Caring*

The notion of "caring" and its relationship with empathy constitutes probably one of the most controversial and animatedly debated topic among empathy researchers. Indeed, we might say that empathy's battlefield is divided between those who argue for the moral relevance of empathy, directly relating it to the notion of care (e.g. Hoffman 2007; Slote 2013; Vignemont and Jacob 2012) and those who regard this view as a commonsensical misunderstanding of the phenomenon and claim that our moral lives would surely be better off without empathy (e.g. Prinz 2011; Bloom 2017, 2018). In between these two main factions there is a large number of theorists who stress both empathy's flaws and virtues, reject univocal, clear-cut judgments about the import of empathy on our moral conduct (e.g. Goldie 2000, Nussbaum 2001), and hold that it can be deployed either for markedly prosocial purposes or explicitly harmful goals.

It should not be a surprise by now, this large variety of disagreement depends on the specific account of empathy various scholars in the field seem to rely on. To make matters even more intricated, the way in which the relationship between empathy and care is spelled out does not merely depend on what account of empathy one deploys, but also on the selective emphasis given by scholars to evidence supporting either the morally positive or negative features of empathy.

Coming back to our list of empathic phenomena we can see that, on some accounts, caring is simply another name for empathy. The notions of concern, sympathy and compassion, enlisted at point (v) in the previous section, seem by definition linked to some form of care towards the other person.

With the other notions, the relationship gets blurrier. Emotional contagion (i) is indeed something that can occur independently of any subsequent manifestation of care: e.g. if I catch your sadness, I might well ruminate on my own personal distress while being indifferent to yours.

The relationship between caring and empathy is even less robust when the latter is merely conceived as accurate mindreading (iii): if empathy is broadly conceived as the capacity to accurately infer the goings-on of somebody else's mind, then there is no clear way in which mindreading should necessarily entail caring.

Affective sharing (ii) as well swings between two extremes: on the one hand, it seems natural to argue that if I come to vicariously share e.g. your suffering, then I might come to care about your suffering, especially if the empathizer's emotional engagement with the target's experience is accompanied by interpersonal understanding (as prescribed by most accounts of empathy as affective sharing). At the same time, though, there is no reason to see how mere affective sharing should *necessarily* conduce to caring for the other person. For instance, it might very well be that caring attitudes are overridden by other motives, that prevail on our prosocial incentives.

Similar considerations can be extended to the case of perspective-taking (iv). By coming to take on the perspective of the other person, we can surely come to appreciate – among other things – their needs and wills from a first-person perspective. Perspective-taking could also result in the empathizer coming to share the emotion the target might be experiencing. Still, there are no reasons that force us to infer that perspective-taking should necessarily lead to caring behaviors or attitudes. It is an indisputable triviality that, in our daily life, we take other people's perspectives for purposes that have less noble motives and outcomes than caring: e.g. we might want to figure out what would be the most appropriate way to deceive someone.[4]

---

[4] One aspect that is very rarely addressed in the literature is the relationship between empathy and caring on an inversed causal relationship: could it be that *caring* for someone might trigger empathic manifestations rather than the other way around? Not many scholars have explicitly addressed this issue (with the notable exception of Bloom 2016), even though, a quick look at our daily experiences seems to immediately validate this suggestion. How many times have we found ourselves wondering what our partner might be thinking or feeling, or responding with great pain to their pain or to imagine ourselves in their circumstances, etc., precisely because we care so much about them? Unfortunately, we have to leave these questions aside for now.

As we have already announced, some influential work produced by leading scholars in the field (e.g. Bloom 2016; Prinz 2011) has brought good arguments to show how empathy and care can diverge. On the one hand, scholars highlight how empathy leads to caring, on the other hand, scholars are committed to proving how empathy and care could dramatically diverge. That is why Michael prefers to talk about "defeasible caring" rather than mere caring: in so doing, he wants to account for all those cases in which an empathic occurrence could be overridden by other reasons not to care (Michael 2014: 170).

So stated, though, it seems like care is something that is still naturally entrenched with empathy, i.e. all things being equal, empathic phenomena seem to spontaneously lead to caring behaviors or dispositions. This natural path from empathy to care can be interrupted only if other overriding forces intervein in the process. Authors like Bloom and Prinz, though, aim to show how reasons not to care can be found in empathy itself rather than in other non-empathic attitudes that can override empathy's prosocial drives. For instance, Bloom points out, among other things, how empathy can be exhausting to undergo and lead to emotional burnouts, so that little is left for caring or helping behaviors (Bloom 2016): indeed, coming to share someone else's suffering can drain the empathizer's energies, so that little is left to the motivation to care about another's misfortune.

However, even skeptics of empathy as a moral guide and therefore, even those authors who would be uncomfortable to see empathy as a royal avenue for caring about others, do not want to deny empathy's capability to lead to prosocial or caring behaviors altogether but merely want to problematize the widespread assumption that empathy is nothing but a good thing and point out to other cognitive and emotional processes that are maybe better suited for morally praiseworthy conduct. No matter whether you are a skeptic or an empathy-enthusiast (or something in between the two, as it probably seems more reasonable), the notion of care and that of empathy do figure prominently together in many works of scholars in the field so that we care can surely figure among those general elements that help us point out the epistemic field within which most accounts of empathy seem to gather.

5. Towards an Account of Empathy

Engaging with Michael's analysis allowed us to individuate some very broad principles that seem to be shared by *most* accounts of empathy. The level of generality was admittedly high, but it was still possible to see how all those principles could be related to our list of empathic phenomena that was previously presented.

In this section, I am finally going to spell out the account of empathy that will be deployed throughout the rest of this work. I will try to illustrate how the present account of empathy relates to the four general principles enlisted in the previous section and other neighbouring phenomena in the social cognition realm. As it should be clear by now, the present account does not aim at being *the* only possible account of empathy. The previous arguments hopefully succeeded in showing the wideness of divergence among various accounts of empathy, so that at this point it would be unrealistically naïve to find a universal definition that works for everybody.

One simple but nonetheless wise suggestion comes from C. Daniel Batson (Batson 2009): to deal with the endemic variety of what people call "empathy", the best thing scholars can do is to make explicit what notion of empathy they are interested in. This prevents undue terminological confusion and enables a meaningful critical engagement with the relevant empathy-related research.

I fully embrace this suggestion. At the same time, I tried to pave the way towards a definite account of empathy by deploying a strategy that is not merely stipulative but locates empathy within a narrower conceptual. This served the double purpose of limiting the degree of stipulation while at the same time providing us with some conceptual tools that would help me to better clarify the details of the account of empathy that will be deployed throughout this work.

The kind of empathy I am interested in is a process of perspective-taking that aims at first-personally imagining somebody else's perspective in certain given circumstances.

This account is naturally related to what we identified as the principle of interpersonal understanding (a). Indeed, by imagining ourselves occupying the perspective of somebody else, we are supposed to gain some understanding about them. We can not only come to understand *what* their occurrent mental states are but also *why* they hold certain mental states and behaving in a certain way. Moreover, by imaginatively modelling their first-personally centred perspective, we can come to appreciate *what it feels like* to be in their shoes.[5]

For what concerns the notion of engagement (b), we can come back to what we already said in the previous section about Jaspers' account of imaginative understanding. In order to first-personally imagine the perspective of somebody else, the empathizer has to transpose herself into the other's perspective so that she can come to appreciate the ways in which things are experienced from a subjective perspective.

[5] This is a crucial aspect on which we will focus more extensively later on.

The principle of engagement and that of similarity (c) are intimately interconnected: the empathizer has to bring about an imaginative representation that is similar enough to the experience undergone by the target. In one word: she has to bring herself to imaginatively undergo an experience that is relevantly similar to the one undergone by the target.

What we called "defeasible caring" (d) is less straightforwardly related to the present account of empathy. As previously stated, there are no straightforward reasons that force us to think that imaginatively engaging with somebody else's perspective should necessarily lead us to care about them, or their well-being, needs, and aims. It is rather easy to conceive of a situation in which we succeed in imagining ourselves in somebody else's shoes without us necessarily "bringing their case to our own bosoms". We can indeed imagine a situation where someone has an interest that irremediably conflicts with our own so that the satisfaction of one of the two interests necessarily imply the suppression of the other. In this case, we can certainly come to appreciate the intimate reasons behind somebody else's interest, still caring exclusively about the satisfaction of our own interest.

However, if we come to imaginatively represent somebody else's first-person perspective, at the very least we become sensitive to her own interests. In other words, by coming to see the world through somebody else's eyes, we can come to appreciate another person's reasons that *speak for* the kind of interests or purposes she has. That seems to imply the presence of a minimal degree of caring: through empathy we are made sensitive towards what is valuable or choiceworthy from somebody else's perspective.

Of course, this does not rule out the possibility – envisioned by our previous example – that the empathizer might hold other reasons *not* to care about the target: no matter how successfully we manage to model her perspective, we might still have conflicting interest that are incompatible with us caring about theirs. But this is precisely the reason why Michael talked about defeasible caring rather than mere caring. Coming to see the world through somebody else's eyes makes us sensitive to what seems of value from their perspective and this seems to imply at least a minimal degree of caring.[6]

---

[6] With these considerations I am not embracing a philosophical view according to which caring about someone amounts to conform to what the target of our care explicitly wants from his first-person point of view. As insightfully noticed by **Darwall (1997: 262),** the notion of care does not require "either i) that what it makes sense for a person to desire and seek will contribute to his well-being nor ii) that whatever will advance his good is something he should rationally seek, even prima facie". While I agree with Darwall's observations, I do also believe that some central instances of care dramatically depend on a subject's capability to take into account, and to some extent endorse, some – or most – of what

As we, saw, this "minimal" care can be easily overridden by other conflicting reasons we might have not to care. Therefore, this account of empathy should be treated with extreme caution when it comes to assess its possible prosocial contribution. In what follows, I will tend to consider it as generally neutral moral force, one that could be in principle be efficiently deployed for prosocial or explicitly harmful purposes.[7]

At the same time, I am willing to acknowledge its full potential as a tool to foster and promote moral virtues such as tolerance and mutual respect. Moreover, given the high level of interpersonal understanding that it can enable us to achieve, empathy can sometimes be an invaluably powerful tool to strengthen social bonds (Carruthers 2020).

I will come back to some of these issues more extensively later on in the present work, when I will deal with issues related to the possibility of learning through empathy and the potential it has to change us as individuals.

Tackling all these issues at the moment though, would certainly be premature. We still need to make a couple of more steps before we can delve into the topic of what empathy can and cannot do.

6. Empathy and knowing what it's like: Introducing phenomenal insight

So far, I have introduced some general characteristics of the account of empathy that will be privileged throughout this work[8] and to which the various philosophical issues that will be addressed are more directly linked to.

As it stands, empathy is a device for interpersonal understanding, specifically one that allows us (when successfully deployed, of course) to appreciate somebody else's state as she experiences it from her first-person perspective. This act is taken to be achieved through a process of imaginative transposition whereby we re-centre our subjective standpoint onto the other's first-person perspective. This process is supposed to endow us with a grasp on "what it feels like" to occupy the other person's shoes in a given circumstance.

---

the other person holds as worthy for himself. Empathy can immensely contribute to developing that kind of sensitivity towards another person's first-personally held desires, purposes, etc. Unfortunately, going into the details of such a fascinating topic is something that falls outside the scope of the present chapter.

[7] In this context I am using the term "moral" in an admittedly broad and naïve sense, it merely indicates any kind of behavior or attitude that could be considered as contributing to the "good" of our social lives.

[8] From now on, unless otherwise specified, simply "empathy".

I will label this particular achievement of empathy "phenomenal insight", i.e. empathy's capability to make us inhabit an alien point of view and let us appreciate what it feels like to be in it. Phenomenal insight individuates one of empathy's specific contribution to human social cognition: it tells us what *kind* of understanding it is supposed to deliver. Also, phenomenal insight is what allows us to individuate and distinguish empathy from other closely related forms of interpersonal understanding: it does not merely amount to the capacity to accurately "label" somebody else's mental states, but rather requires us to be able to represent them from a subjective point of view. This requires us to be more specific about how to conceive of the kind of imagination deployed by empathy in the process of gaining phenomenal insight into the other person's perspective.

In what follows, I am going to specify some of the fundamental features of phenomenal insight and its specific contribution to interpersonal understanding that I take it to deliver. This will also require us to delve into some of the features of the kind of imagination on which empathy relies.

6.1. Phenomenal insight: Two Preliminary Clarifications

With the notion of phenomenal insight, I aim in this work to capture a rich sense in which a subject looks at the world through a first-person perspective or specific point of view.

The adjective "phenomenal", therefore, needs to be conceived in a considerably broad sense in order to capture the whole fact of subjective experience. Therefore, phenomenal insight is not merely concerned with the representation of what philosophers have called "*qualia*", i.e. the qualitative character of mental states. Representing a subjectively conscious point of view via phenomenal insight does certainly involve also the conjuring up of the relevant qualitative features but does not reduce to that.

Therefore, in the context of this paper, phenomenology will never be a matter of *pure* internal vestiges of experience but will be conceived as intrinsically *impure*, i.e. constantly embedded within the larger frame of non-qualitative internal and external circumstances experiencing subjects find themselves in.[9] For instance, phenomenal insight into the feeling of resentment for a lover's betrayal, will not be a matter of rehearsing the mere internal *feel* of resentment but would require us to *embed* this feeling within the concrete situation of betrayal it relates to.

---

[9] For similar considerations, check (Campbell 2015; Paul 2015a). The phrase "impure phenomenology" is borrowed from (Boyle 2020) who deploys it to account for the specific phenomenological features of episodic memory.

Another disambiguation that needs to be put in place can be spelled out with reference to some remarks made by Stephen Darwall (Darwall 1998). He writes:

> Consider the difference between the instructions: (a) imagine what someone would feel if he were to lose his only child, and (b) imagine what it would be like for that person to feel that way. Complying with (a) involves simulating someone in the imagined circumstances in order to identify what feelings the situation would apparently warrant when so viewed. It need involve no attention at all to what it would be like for the person to have those feelings or suffer that loss. To comply with the second request, however, one would have to simulate, not just the person *with* the relevant feelings, but someone *conscious of* his feelings, their phenomenological textures, and relevance for his life. (Darwall 1997: 270)

Darwall is drawing a distinction between two related but nevertheless distinguished processes: one (a) that seems close to our account of empathy, i.e. imagining a certain situation from somebody else's subjective standpoint, and another (b) that amounts to what we called phenomenal insight.

But how can these two things come apart, if we took them to be two faces of the same coin? As far as I can make sense of Darwall's intuitions, (b) does not require what we referred to as the similarity condition, i.e. a representation that actually matches the way in which the *other* experiences a certain situation from her point of view. The process (b) described by Darwall is basically a way of imaginatively *engaging* with the other's experience in order to work out a self-conscious perspective of her situation, rather than an imaginative replica of her point of view. We could say that (b) represents what it feels like for the other person to undergo a certain experience *were she to be introspectively conscious about it*.

Phenomenal insight, as I will deploy it in the present work, does not require such introspective conscious access to one's experience. It is undeniably true that, in some circumstances, the phenomenological features of certain experiences do entail self-conscious representations that are eminently about *me* and what it feels like *for me* to undergo such and such experiences. This kind of self-consciousness is likely to be instantiated e.g. whenever we answer questions such as "how do you feel about it?" which seem to entail that we explicitly bring ourselves in the evaluation of what an experience felt like *for us*.

Darwall's example is illuminating on this particular point:

> Someone who has lost a child might be so consumed by the loss that he is unable even to think about what living with it is like *for him*. […] Or perhaps the loss is so devastating that he denies it, thinking and acting as though the child were still alive. (Darwall 1997: 271)

As far as I understand it, Darwall shows that we can perfectly imagine a father who has lost his child and that is not conscious, from his subjective standpoint, of what that experience must feel like for him, i.e. he might be fully oblivious of *how* that experience explicitly relates to him and what it means *for him* to undergo it.

At the same time, though, the father in the example is undeniably undergoing certain kinds of experiences from a first-person perspective, even though he is not taking *himself* as the main target of his focus. The fact that we do not explicitly take into account *ourselves* when we go about *our* lives, does not entail that there is nothing that it feels like for us to undergoing these experiences from *our* perspective. To put it roughly, there are experiences that do feel like something, even though we are not explicitly representing that *we* are undergoing them. Similarly, it might well be that the other person is not introspectively conscious of how things might be like for *her*, while at the same time relating in a first-personal way to them. Therefore, it certainly makes sense to talk about "what it feels like" to be in the other person's shoes when all we have to imagine is the first-personal way to relate to certain contents of experience even if these do not include a conscious self-representation.

To better illustrate this point, it might be useful to look for a moment to Nagel's famous paper "What is it like to be a bat?" (Nagel 1974). In this case, we have good reasons to believe that bats are not self-conscious in any robust way, i.e. it is very unlikely that they can entertain representations that explicitly take into account how certain experiences feel *for them*. At the same time – and this is what made Nagel's paper so influential –, it certainly makes sense to wonder about what it could feel like to be a bat. If that is the case, then in order for experiences to *feel like* something for a subject experiencing them it is not necessary that subjects be conscious about how these experiences feel *for them*. [10]

Bringing the discussion back to humans, I can be experiencing the fatigue of climbing a mountain, and there might be something that it feels like for me to undergo that experience, without me being explicitly self-referencing that experience, i.e. without me having a representation of *myself* undergoing the experience. I might be totally focused on the steepness of the mountain and how

[10] Certain authors would strongly deny altogether the fact that bats have experiences that *feel like* something. These authors are usually aligned with a theory of phenomenal consciousness as globally broadcast nonconceptual contents (Carruthers 2020). It might well turn out that these scholars are right and that, indeed, there is no matter of the fact about bats' phenomenal consciousness. Still, the intuition delivered by my example would hold true. Indeed, I was simply trying to point out that it sounds perfectly reasonable to talk about phenomenological qualities of experiences, even though no robust self-representation comes into play. This is left untouched by the fact that, in the future, it might actually turn out that bats do not have phenomenal consciousness. The fact that they do not have it does not indeed depend on the fact that they do not have any robust self-representation but to the fact they do not have globally broadcast non-conceptual content.

terribly far the top looks from where I am, without entertaining any representation of how it feels like *for me* to undergo those experiences.

In a nutshell: the phenomenal qualities of my experience, while depending on my specific first-personal access to them and while being constantly self-ascribed as features of experiences that *I* am undergoing, can be utterly outwardly directed, i.e. they do not have always to be accompanied by an *explicit* self-representation. If I am tasting coconut, I might not entertain representation that overtly relates the tasting of coconut to what that experience feels like for me: I might simply be enjoying the fresh delight of tasting coconut. I am of course still the one to which that experience refers, but only as the implicit subjective centre in which that experience unfolds, not as the explicit target of my attention.

Some remarks made by François Recanati (Recanati 2007: 203- 204)) turn out to be useful to further clarify my point.[11] There are two ways in which I can entertain *de se* imaginings, i.e. first-personal imaginings: explicit *de se* imaginings and implicit *de se* imaginings. In the first case, the self figures in the content of what is imagined, in the second case the self is only implicitly represented. Differently from explicit *de se* imaginings, with implicit *de se* imaginings, the self does not figure as one of the constituents of what is imagined. The only thing that figures are features that are self-ascribed, i.e. related to a self that is only implicitly present as the first-personal centre of the experience.

These remarks can help us make sense of Darwall's example. The father in the example does not have to consciously work out what all this might mean *for him*, he might simply feel terrible. Still, and in this consists my remark, his experiences *do feel* like something from a first-person perspective, i.e. it still makes sense to wonder "what it might feel like" to be in his shoes. In this case, phenomenal insight does not coincide with the interpretation Darwall has given to it. In the context of the present work, phenomenal insight is simply concerned with the re-creation of the perspectival nature of experiences as they are experienced from the phenomenally conscious point of view of a subject.

6.2. Imagining another's perspective via belief-like imagination

Having taken care of some preliminary concerns of what I mean by phenomenal insight, it is now time to get deeper into the kind of psychological mechanism that underlies it, namely the imagination.

---

[11] Recanati is interested in issues that do not concern us here (at the moment). Therefore, my strategy will consist in extracting his observation from their original context and instrumentally re-locate them in the context of our present discussion.

It is common, in contemporary debates about imagination, to distinguish between different *kinds* of imagination.

One relevant type of imagination, widely deployed in different contexts of our everyday life, is called belief-like imagination, also known as propositional imagination (Nichols 2020).[12] Belief-like imagination is merely concerned with the truth-values of given propositions, i.e. it mainly consists in *taking* a proposition to be true. For instance, if I were to propositionally imagine that "pigs have wings", all I would have to do would simply be to imaginatively take this proposition to be true.

Now we can ask, what are the epistemic upshots of such a cognitive process? Among many others, one important outcome is that subjects are endowed with the power to explore the logical space of possibility thanks to the expanded epistemic landscape opened up by the imagination. Coming back to the winged pigs' example, thanks to belief-like imagination we would be in the position to make informed inferences about a possible world in which pigs have wings: i.e. it is likely that pigs would be kept in aviaries rather than in paddocks or – even better – pigs would be free from human captivity altogether and spend their lives blissfully flying in the sky.

Taking a less exotic example, if I were to belief-like imagine that "I left my laptop outside during a storm", then, again, I would be in the position to draw informed inferences about this possible state of affairs such as "it is very likely that my laptop would be unusable".

As aptly noted by Gregory Currie (Currie 2004: 174)**,** when we engage in this kind of imaginative process: "we hold an assumption in thought, consider its consequences and its relations to what else we think is true, and covert the assumption into a belief", i.e. we draw inferences on the basis of such an operation.

One observation is now in order. We can plainly imagine the above-mentioned propositions without having to model any concrete experiential features that might be linked to them. That is, I can imagine that my laptop is unusable after being exposed to a storm, without having to imagine it under the drops of rain. This means that we can perfectly engage in exercises of belief-like imagination without having to imagine any relevant experiential features that might be linked to the content of the

[12] Throughout the present work I will mainly use propositional imagination and belief-like imagination interchangeably. It should be noted that propositional imagination could be taken to constitute a broader category than belief-like imagination. I can indeed imagine a variety of propositional states and it is not at all obvious that all these states should be imagined by means of belief-like imagination. I will partly come back to this issue when I will discuss the issue of imagining desires. For now, I will leave this controversy aside and align to the widespread tendency in contemporary debates to equate belief-like and propositional imagination.

propositions imagined, i.e. we do not have to imagine *experiencing* these facts. Using our terminology, the deployment of belief-like imagination is orthogonal to phenomenal insight.

We can now see how these considerations apply to the case of mutual understanding. We can indeed see how propositional imagination could be fruitfully employed to work out the goings-on of other people's minds and to predict or explain their behaviour in given circumstances.

It is often said that in order to perform this imaginative act of perspective-taking, we need to imaginatively "take on" some of the relevant mental states held by the target. The two canonical instances of mental states often analysed in the specialised literature are beliefs and desires. Suppose that we see a friend going to the fridge and grab a bottle of beer. One way of making sense of his behaviour consists in imaginatively entertaining a relevant belief-desire couple and see how it fits with the behaviour displayed.

Interestingly enough, such a process can be in principle accomplished also by means of what we called propositional imagination: we can simply deploy our imagination to take certain propositional attitudes to be instantiated in the context of the other person's mind. In light of these propositions, we can satisfactorily explain the other person's behaviour: our friend went to the fridge and grabbed a bottle of beer *because* the desire to drink a beer and the belief that there was a beer in the fridge hold true for her. This process allows us not only to explain but also to predict another person's behaviour that is likely to occur *in light* of the propositional attitudes that we take to hold.

Before moving forward two brief clarifications are in order:

i) The debate concerning what kind of knowledge is responsible for our capability to explain and/or predict other people's behaviours on the basis of some propositional attitudes that we imagine them to hold. It might well be a folk-theory about human psychology, or it might some other kind of culturally or otherwise embedded knowledge concerning causal and correlational links between mental states and behaviours. These are worries that can be set aside for the moment. What matters is that we can evaluate other people's mental goings-on and behaviours by deploying one relevant kind of imagination, i.e. belief-like imagination.

ii) Those who are familiar with the relevant literature concerning the so-called desire-like imagination, might consider my move to account for imagined desires in terms of belief-like imagination somewhat odd. Indeed, as noted by Currie and Ravenscroft (2002), desires should not be represented by means of belief-like imagination. Imagining a desire, they claim, does not amount to merely belief-like imagining that a person holds a desire

> D, because that would confuse the *character* or intentional structure of a mental states such as desires with their *contents*. In other words, the intentional structure (what Currie and Ravenscroft call the "character") of desires cannot be captured by belief-like imagination. To account for this concern, I can make two observations. First, their considerations are advanced in the context of their theory of re-creative imagination, according to which re-creative imagination re-instantiates imaginative replicas of other non-imaginative mental states. In this sense, to re-creatively imagine a desire would require us to instantiate an imaginative replica of a non-imaginative state of desire. Only in the context of re-creative imagination, then we need to take care not only of the content but also of the character of the relevant mental state. Therefore, only if we are interested in re-creating mental states should we take into account Currie and Ravenscroft's worry. Second, the fact that to imaginatively re-create a desire (in Currie and Ravenscroft's sense) does not involve belief-like imagination does not entail that it is impossible to belief-like imagine desires. Indeed, as we saw, belief-like imagining desires can allow – among other things – subjects to make explanations and predictions about other people's mental states and behaviours.

Now, if we are willing to grant that belief-like imagination can be deployed to assess the range of possible mental states agents might hold with the purpose of making predictions or explanations about their minds and behaviours, we can also see how this can differ from other kinds of imagination that need to be in place in order for phenomenal insight to be achieved. Indeed, as we already noted in the two examples at the beginning of this section, belief-like imagination is something that can be achieved without any relevant experiential features being conjured up in the imaginative process. Therefore, belief-like imagination seems, by definition, incapable of providing us with an understanding of what it might feel like to be in the other person's shoes.

6.3. Phenomenal Insight and Imagination: Experiential Embedding

One observation is now in order. From what I just said in the previous section, it might seem that I am indirectly espousing the so-called theory of "cognitive phenomenology" (Bayne and Montague 2011; Pitt 2004). By claiming that we can propositionally imagine beliefs and desires without any relevant experiential features being instantiated, it might seem that I am willing to grant the possibility that these mental states have a distinctive phenomenology.

In principle, I do not have anything particularly strong to assert against this view: there seems to be something intuitively true about the claim that at least *some* conscious occurrences of beliefs and desires can be accompanied by a distinctive phenomenology. I am not particularly afraid of the

possibility that arguments in favour of some kind of cognitive phenomenology and, specifically, about the distinctive phenomenology of attitudinal states such as beliefs and desires might turn out not only to sound intuitively true but also philosophically solid.

At the same time, though, I do not have to commit to any theoretically heavy-weight assumptions about the phenomenology of cognition in order for my considerations to make sense. What we need to grant is not the existence of a distinctive experiential marker of mental states such as beliefs and desires that could *per se* set them apart from other kinds of mental states such as e.g. perceptions or emotions. All we need is to acknowledge is the fact that, in the context of an individual's conscious life, these mental states are always *embedded* in a peculiar perspective, i.e. in a particular experiential setting that has the subject as its conscious centre. In this sense, to conjure up beliefs and desires' *experiential features* would simply require us to consider them as relating to the subject's specific first-person perspective into which they are embedded, rather than as mere propositions that can be evaluated purely in terms of their truth-values.

However, it should be acknowledged that propositional attitudes such as beliefs and desires do not *necessarily* come with a distinctive phenomenology. Indeed, their content seems to be satisfactorily captured in propositional terms that can be plainly dissociable by any relevant experiential property. Coming back to our previous point, it sounds perfectly sound to claim that an occurrent belief that "democracy is the best possible form of government" is not essentially linked to any relevant phenomenological feature. This belief seems to be fully captured by the proposition expressing it in the natural language.

Other beliefs, though, seem to be more intimately related to some phenomenological features. As aptly noted by Amy Kind (Kind 2001: 93):

> Some of our demonstrative beliefs about colors might be thought to provide plausible examples: suppose that, while looking at my friend's dress, I were to have the belief "That is an ugly shade of orange." It is hard to see how this belief could be occurrent without involving a state with a qualitative feel, that is, a sensation of orange.

However, Kind warns us that the phenomenological features that might be related to beliefs are not to be ascribed to the belief state itself but rather to the *content* conveyed by the belief. Even though I tend to agree with Kind's critical observation, it is still possible to claim that, in some cases – e.g. demonstrative beliefs – certain beliefs need to come with certain experiential features conveyed by their contents in order to be accurately imagined. Only if we embrace a "compartmentalized" and quite abstract picture of the mind could we be fine with the idea of empathic understanding as being

conveyed by mere propositions that aim to describe a subject's first-person perspective in a given context.

Most of our beliefs (ad desires) are *embedded* within the larger context of our mental life. When I empathize with certain attitudinal states held by a subject, I need to get a grasp on the role they play in the broader context of his lived experience. For instance, if I want to empathize with someone who believes that "democracy is the best possible form of government", it is not enough to merely grasp that single proposition describing his mental state. The nature of that occurrent belief, indeed, is not fully captured by the proposition describing it, but it is linked to other important facts about the psychology of the person, e.g. their political preferences, commitments, values, aspirations, etc. We might get the illusion that beliefs or desires amount to nothing but the proposition that expresses them. But actually, when it comes to a subject's lived experience, these states have a much broader meaning: they constitute a specific *take* on the world. And here is where talks about the what-is-likeness of certain attitudinal states start to make sense. When we embed mental states within the larger context of people's subjective experience, we see how relevant and appropriate it becomes to ask question such as "what is it like to hold such and such beliefs and desires?".

Especially if we are willing to embrace, as we made clear in the previous section, a conception of phenomenal insight that is not merely concerned with the qualia-like characters of mental states, we have to be ready to admit that attitudinal states such as beliefs and desires *do* feel like something. With this, I do not want to claim that *every* occurrence of beliefs and desires will always have a distinctive phenomenological profile, or at least one that is robust enough for it to make it a worthwhile enquiry for empathic exploration. I am perfectly fine with the claim that there is no particular way in which the belief that "2 + 2 = 4" feels like for the subjects holding it.

In order to get an intuitive grasp on this point, it could be useful to focus on the classic distinction between semantic memory and episodic memory.

As it is widely accepted in the science and philosophy of memory, semantic memory is the retained knowledge about certain facts or states of affairs that are taken to obtain (or to have once obtained) in the world. For example, as part of my semantic memory figure states that can be captured by propositions of this sort: "Rome is the capital of Italy", "my previous flat used to be in Rue César Franck, Paris" etc.

Differently from semantic memory, episodic memory has been taken to involve a kind of "mental time travel". Subjects episodically recollecting an experience, "travel back in time" through memory

to consciously represent *experiencing* episodes of their own past. Ideally, this process is taken to re-instantiate some of the relevant objective and subjective features of the event retrieved.

It is interesting to notice that semantic memory as well can be about some facts that we experienced in the past or, as some scholars would put it, the contents of semantic memory can be *de se*. For instance, I can retain as semantic memory the information that I once entered the Sistine Chapel and yet be utterly incapable of conjuring up any relevant experiential features of that visit – I might not recall the subjects painted on the ceiling and what it felt like to witness such a powerful display of human creativity. In one word, I can store as semantic memory *de se* facts, without being able to episodically recollect those facts as I happened to experience them, i.e. without being capable of *embedding* these facts in the actual first-person perspective I was occupying at that time.

As a thought experiment, imagine that we were offered the possibility to visit for free a city we always wanted to see. Imagine that, as a contract clause, though, we can only retain semantic memories of our trip: we would know for sure that we were there, without being able to recollect the *experience* of actually having been there. It seems that this clause would play a huge role in our decision whether to undertake this trip or not and – so intuitions seem to tell us – the experience would lose much of its value.

This striking difference between semantic and episodic memory can be easily transposed to the context of belief-like imagination and empathy. The difference between semantically remembering something and episodically recollecting an experience can make us appreciate the difference between two ways of accessing the goings-on of other people's mind: one thing is to merely be in the position to formulate propositions that accurately describe certain relevant features of the other person's perspective, another thing is to empathically explore that perspective.

6.4. A Brief Clarification

In the above considerations, I tended to focus almost exclusively on propositional attitudes such as beliefs and desires and highlighted how it is still possible to empathize with – and therefore achieve phenomenal insight into – these states. I focused on them because the relationship between empathy, phenomenal insight and beliefs and desires seemed to be less straightforwardly intuitive and, therefore, in need of a more urgent justification.

However, this should not give us the wrong impression that empathy is exclusively concerned with the imaginative understanding of propositional attitudes. Indeed, phenomenal insight and empathy seem to be much more easily related to states that have a rich phenomenological profile such as e.g. affective and emotional experiences. For instance, it makes perfectly sense to empathize with the

immense joy of someone finally reuniting with her best friends after years they spent apart from each other. Or, coming back to Darwall's example, there is plenty of work for empathy when we are facing with the tragic experience of a father losing his son. In these cases, it makes perfectly sense to conceive of empathy's phenomenal insight as having to do with the imaginative recreation of some of the relevant experiential features of those states.

In cases where feelings are involved, empathy and phenomenal insight seem to be even more aptly involved. Moreover, as insightfully noted – among others – by Gregory Currie (Currie 2004) and Kendall Walton (K. Walton 2015) – most of our folk-intuitions about empathy are concerned with the sharing or imagination of centrally affective and emotional experiences. Even though – as Currie and Walton as well point out – we do not have any principled reason to reduce the empathic scope to the experience that centrally involve feelings, we do not have to forget that in many occasions empathic understanding of feelings is mainly what we are interested in.

Our previous considerations, while focusing at length on the empathic understanding of propositional attitudes might have given the false impression that beliefs and desires are everything empathy is concerned with.

In Chapter II, I will come back to the stimulating topic of empathically understanding emotions. As we will see, phenomenal insight into affective states turn out to be an indispensable ingredient to empathically understand certain forms of agency and the empathic representation of emotional/affective states can constitute a valuable epistemic goal of its own in our practices of mutual understanding.

For now, though, let us stop here.

7. Do We Really Need the First-Person Perspective?

In articulating some of the relevant features of empathy, I recurrently used expressions such as "first-person perspective", "subjective experience", "subjective point of view", etc. With these, I aimed to capture the specific epistemic goal of empathy.

Talks of perspectivity are useful and commonly deployed in empathy debates: nearly everybody interested in empathy as a form of *perspective-taking* cannot, by definition, omit referring to perspectives.

Not everybody in the empathy debate, though, feels comfortable in using a perspective-laden terminology. Hermann Cappelen and Josh Dever have come to be known for expressing harsh

criticism against the philosophical tendency to deploy terms such as “first-person perspective”, “essential indexical”, “first-personal modes of presentation”, etc. (Cappelen and Dever 2013). According to these authors, the aforementioned notions have mistakenly taken by an influential and large group of philosophers to capture some fundamentally important philosophical issues, whereas – so they claim – all there is to be found beneath the surface of these terms is, at best, something that can be better explained by a perspective-free terminology or, at worst, philosophical non-sense riddled with ambiguities and confusion.

I do not aim here to investigate and critically assess their general view on this issue, because it would require a work of its own and would definitely lead us off-topic. I will rather address one of their papers (Cappelen and Dever 2017) in which they explicitly extend their account to the issue of empathy. They claim that talks of first-person perspectives and the like are of no use for a thorough and philosophically sound account of empathy. As I largely committed to the terminology Cappellen and Dever (henceforth C&D) so forcefully demonize, it is useful to take into account their challenge so that we can assess whether it poses a serious threat to my line of argumentation.

7.1. Cappelen & Dever’s Critique

C&D’s critique against the supposed importance of the first-person perspective in the context of empathy is to be located within the larger context of a critical engagement with some relevant work by Laurie Paul on empathy and transformative experiences (Paul 2014, 2017). Paul, indeed, figures among those authors who explicitly highlighted the importance of a first-person point of view in empathy.

In this section, I am going to leave aside the issue of transformative experiences and focus exclusively on the topic of empathy. I shall come back at length on transformative experiences and empathy in *Chapter V***.**

C&D lay down very clearly their goals at the beginning of their paper:

> We are skeptics about the first person point of view, but we are not skeptics about empathy. In this paper we thus investigate these claims of connections among these notions. We conclude:
>
> 1) that the philosophically important property of empathy has nothing to do with any ‘first person point of view’
> 2) that talk of ‘subjectivity’ in discussions of empathy (and in many other philosophical areas) equivocates between an *experiential* notion that does play an important role in empathy but which is not implicated in

> anything distinctively first-personal or perspectival and a *perspectival* notion which we find philosophically objectionable, but which plays no important role in empathy (Cappellen & Dever 2017: 316-317)[13]

In what follows I will try to show how C&D fail to meet their goals. Indeed, either their critique of the first-person perspective and empathy amounts to a purely verbal dispute (which would make *their critique* against the philosophical unimportance of the first-person philosophically shallow) or, if we want to take their account of empathy at face-value, we need to rehabilitate the notion of perspectives to better understand its fundamental features.

As a start, it is interesting to notice how C&D's conception of empathy is pleasantly in tune with the account I have been sketching so far in the previous sections. They offer the following definition:

> To empathize with someone is to build a model of a person's cognitive and emotional state, for the purposes of understanding, explaining, and predicting that person. (Cappellen & Dever 2017: 320)

They immediately point out that empathy's function is to be conceived of as primarily epistemic, i.e. as concerning the kind of knowledge and understanding that we can achieve about the other person. They specify the necessity of building sufficiently complex and fine-grained models about the other person's mental states so that empathizers get at what it is like for the other person to undergo a certain kind of experience.

To better illustrate this point, they take the example of depression:

> To empathize with the depressed person, we need to get at what it's like for them to be depressed. What does that amount to? It amounts to *having a grip* on lots of the small details of depression. We need to *understand*, for example, that there's a constant flatness of affect, that things don't excite them. We need to understand that they constantly question why other people say nice things about them. (Cappelen & Dever 2017: 320)

It is clear from this description that the kind of modelling that they have in mind is something that needs to get at many relevant features of another subject's experience so that the empathizer could be in the position to formulate "sufficiently fine-grained predictions about what the subject will be thinking and doing in a particular situation" (Cappelen & Dever 2017: 23).

C&D share my scepticism towards the deployment of purely propositional imagination when trying to empathically understand the state the other person is in. As they say, imagination (in the relevant

---

[13] C&D mention also a third important goal, namely showing that "the lessons thus learned from the discussion of empathy extend to Paul's views on transformative experience, which are also *best* articulated without appeal to the 'first person point of view' (Cappelen & Dever 2017: 317). I deliberately omitted this last point because, as said, I will be focusing exclusively on empathy and therefore leave aside C&D's critical assessment of transformative experience.

sense of merely "imagining that" or "supposing that", hence in the sense of "propositional imagination") is indeed too "cheap" to drive empathy.

As I pointed out extensively in the previous sections, it is possible to propositionally imagine certain mental states instantiated by another subject without our imaginings necessarily instantiating a full appreciation to what it is like to be the other person in a given circumstance.

So, at its core, C&D's account of empathy presents a good share of the relevant features I take empathy to incorporate. Divergences emerge when they then claim to spell out their account in a completely perspective-free language.

I address this issue in the next section.

7.2. Why Cappelen and Dever's critique misses its goal

It should be noticed that C&D are not against *every* notion of perspective. Even for them, perspectivality can be legitimately deployed in certain contexts to account for some basic features of cognition. Specifically, they claim to be perfectly comfortable with spatial and cognitive notions of perspective. For what concerns the first notion, talking about perspectives is needed to account for the fact that our visual field is framed within a specific point of view. In this very moment, the objects and features that populate my visual field are apprehended from a specific location: I can see e.g. the front of my laptop while writing this dissertation. If I were to stand up and walk around my room, different objects and different features will show up in my visual field: I could see e.g. the back of my laptop, or my acoustic guitar laid on the bed behind me. In one word, we always see the world from a particular location – or perspective.

Cognitive perspectives instead are a matter of certain beliefs a subject holds on a topic (Cappelen and Dever 2017: 317). In this sense, having two different perspectives on e.g. the next presidential election would mean that A believes that candidate-B is the most qualified for being president, whereas C believes that candidate-D is. A and C hold two different beliefs, hence two different perspectives on the topic of the next presidential elections.

First of all, I have to point out that C&D's notion of cognitive perspective is far too restrictive. They reduce it to the mere fact of diverging beliefs held by different subjects. However, it should be noted that a notion of cognitive perspectives could and probably should be extended to other classes of mental states. Two individuals can, for instance, hold two different *desires* about the same thing. Suppose that I am with a friend watching a football game, I support Juventus, whereas he supports Manchester United. As a consequence of this, we have two radically diverging desires. I guess it is

safe to infer from this story that we hold two different perspectives on the same event, i.e. the football match.

We can go even further and argue that two individuals can have two different *emotional* perspectives on the same thing. For instance, a dog-lover and a dog-hater have two radically different emotional takes on the same subject, i.e. dogs.

Let us be charitable with C&D and, for the sake of the argument, imagine that they would not be bothered too much by the inclusion of other classes of mental states such as (at least) emotions and desirers in their notion of cognitive perspective and that they could find a way to successfully accommodate such states in order to fit with their philosophically "innocuous" notion of perspective.[14]

Still, more robust points of criticism can be raised against the ways in which they argue for a perspective-free description of empathy.

As already stated, C&D add on the notion of model building the necessary to get a grasp on what it is like to be in the other person's state. Why is this necessary? It is necessary because, besides enabling the empathizer to make sufficiently fine-grained predictions and explanations about the other person's mind and behaviour, it also enables them to *understand* their targets, i.e. to get at what it is like to be them.

Similarly to what we did in the *Section 4.1.*, we can imagine a sophisticated machine that, by processing an impressive amount of biometric data and monitoring the neural activity of a given subject, is capable of predicting their mental states and behaviour with a level of accuracy that strikingly outpowers that of the average human mindreader. We can also imagine that this machine works "in reverse" so that, besides prediction, it is also capable of reconstructing an accurate causal history that *explains* the subject's displayed behaviours and previously held mental states.

Even if we grant this fancy machine all this mindreading power, we could still find a way to safely argue that it could lack empathy. The reason why this would be the case is to be found in what we might call the "experiential requirement": the mindreader needs to have a grasp on how certain experiences feel like for subjects undergoing them in order to count as an empathizer.

[14] It should be acknowledged anyway that this move would still prove the pervasiveness of perspectivality in the context of human cognition. Perspectives would turn out to be a fundamental dimension of our mental lives and this is likely to shed more than some reasonable concern on C&D's philosophical aims of largely getting rid of the notion of perspectives.

Now, C&D claim that this experiential requirement can be met without having to mention at any point the first-person perspective. We can interpret this move in two ways: either C&D aim to prove that certain relevant features of empathy *can* be expressed without using any first-person terminology, or that they *should* be articulated without deploying the first-person. Let us examine both intepretations in turn.

In the first case, C&D's critique would amount to nothing more than a mere rhetoric exercise in which the authors prove that the term "first-person perspective" (and cognate expressions) can be successfully rephrased deploying other locutions and formulations.

They indeed point out on several occasions throughout their paper how their "deflationary" account of empathy can be spelled out without mentioning perspectivality. Still, it is very difficult to get rid of the impression that they just deliberately omitted, somewhat deceitfully, those expressions that contain terms such as "first-person perspective" and the like while pointing at a bundle of issues that could still be successfully captured by the language of perspectives. Therefore, if C&'s arguments boil down to the mere attempt to show that it is possible to make a philosophically thorough analysis of empathy without mentioning perspectives, then we are left with no sufficient reason to embrace their strategy and avoid talks of perspectives. If that was the case, their argument would, at best, count as a philosophically negligible rhetoric exercise and could, as such, be safely ignored by anyone who aims to reason about empathy in terms of perspectives.

A more charitable interpretation of C&D's criticism would start by acknowledging their intent to claim that talks of first-person perspectives *should* be avoided in order to build a philosophically sound account of empathy. Indeed, they state it very clearly: "the first-person-talk is just garnish" (Cappelen and Dever 2017: 332). According to them, there is no substantial contribution made by the first-person in spelling out empathy's relevant features. Moreover, garnish entails a philosophical risk, namely that it can be mistaken for substance, therefore leading the conceptual analysis astray. C&D could be therefore claiming, in a somewhat Wittgensteinian spirit, that avoiding ambiguous terminology counts as a philosophical virtue. Therefore, "virtuous" philosophers interested in empathy, *should* avoid talks about perspectivality in order to minimize the risk of ambiguity and undue confusion. This interpretation makes C&D's goal more philosophically relevant, but certainly no less problematic.

In order to articulate more intuitively their "deflationary" account of empathy, C&D ask us to imagine a standard empathic dynamic and see whether there is a real philosophical need for "perspectivality" in order to satisfactorily articulate the essential features of the process:

> Let's imagine, insofar as this is possible, a person – call her Third – who is doing everything third-personally. […] Third can certainly 'try on' the beliefs and attitudes of the person she is trying to empathize with [..]: she can build a partial cognitive model of what it would be to have those experiences and beliefs. What is she lacking? […]
>
> There is a clear and non-mysterious sense in which Third is doing things from her perspective: she, Third, is the one building the model of beliefs and experiences of someone else. She is the one doing it so, in a very clear and non-mysterious sense, it's done from Third's perspective (in the sense that it is Third building the model). What's missing? So far we have not introduced the requirement that Third must present Second's experiences in a 'subjective' way – using a 'subjective' mode of presentation. We have, however, a way to capture at least an aspect of that idea: It's Third who is engaging with E (i.e. with Second's experiences) – she is the one doing the cognitive modeling of E. What more should we ask of Third? […]
>
> [T]he modeling she [Third] engages in enables her to understand what it is like to be Second – it enables her to understand a given situation in many respects as Second would. This is a distinctive sort of information (since it is information about how Second understands a situation – as opposed to Fourth or ...). (Cappelen & Dever 2017: 325-326)

There are at least to concerns one might have in reaction to this passage: the first one is specifically related to their account of empathy, the second one regards their general scepticism towards the first-person.

For what concerns the first concern. C&D do not spell out *how* Third is supposed to have a grip on Second's experience of E. They merely state on several occasions the manifest triviality that Third is the one doing the modelling, i.e. that Third is the cognitive agent of the empathic process. From C&D's analysis, we are left with the bare statement that Third needs to be in the position to build sufficiently sophisticated models about Second's experience of E *and* to get at what it is like for Second to undergo E.

However, we might want to ask: *how* is Third supposed to get at the phenomenology of E? C&D leave this question largely unanswered. But here is where perspectivality becomes very handy, if not philosophically necessary, to adequately account for the process involved. Indeed, in order to grasp what it feels like to experience E, Third needs to be capable of representing E's relevant features under the "subjective mode of presentation". I really do not see how else it could possible to represent certain relevant phenomenological facts without bringing to mind *the way in which they are experienced by the subjects undergoing them*, i.e. without mentioning the first-person perspective. I can describe the taste of coconut as accurately as possible by providing long, detailed descriptions about its sugary taste, its peculiar texture, etc. Still, these descriptions would not bring me anywhere close to appreciating what it is like to taste coconut if I am not in the position to represent these

features as *experienced* by a subject, i.e. without representing certain relevant facts concerning the first-personal perspective on the taste of coconut.

There is no convincing reason to exclude from an account of empathy the powerful explanatory device of "perspectives": it is at the same time intuitively appealing and philosophically robust enough to capture fundamental features of the processes involved.

For what concerns the second concern. C&D's scepticism towards the first-person seems to deliberately ignore its pivotal role in the context of human agency and cognition. The first-person mode of presentation of certain relevant states of affairs about the world and about oneself is indeed intimately related to the correct functioning of our mind and our optimal interaction with the environment that surrounds us.

The fact that many phenomena involving indexicality *can* be rephrased in non-indexical terms *salva veritate*, does not necessarily entail that indexicality is philosophically shallow, especially for those areas of philosophy that are interested in the *actual* working of the human mind and not merely in the production of sound stories with no validation in the world we happen to inhabit.

It is surely possible to rephrase the proposition "there is a lion behind the bush" by linking every letter of the proposition to a singular verse of Homer's *Iliad*, so that the "t" of the world "there" corresponds to the first verse, the "h" to the second, and so on and so forth. If one were to possess the interpretative key to map the first formulation on the second, the same content could be conveyed *salva veritate*. However, there would be an immense difference at a psychological level in terms of how efficiently we would apprehend the content conveyed by that proposition if we were to read it as it stands or via the Iliad-mediated rephrasing of it. The way in which a certain meaning is represented matters if the focus of our inquiry is on how the mind works and not merely how logically conceivable minds could work in certain possible worlds. Moral: mental contents that are indexically represented do not always play the same psychological role as their non-indexical counterparts.

We might have an even deeper reason of concern. In certain contexts, e.g. agency, indexical thoughts seem to play an *irreplaceable* psychological role.

Suppose that I entertain the mental content expressed by the following proposition "I am about to be attacked by a bear". This self-ascribed proposition, indexically related to *me*, is likely to have a considerable impact on my subsequent actions, e.g. I might start running away from the bear.

As C&D pointed out (Cappelen & Dever 2013), it is possible to rephrase the above proposition in non-indexical terms *salva veritate*: "I am about to be attacked by a bear" conveys the same content

as "MI is about to be attacked by a bear", i.e. *at the semantic level*, nothing is lost in the transition from the first to the second formulation. However, the way in which that content is presented *at the psychological level* matters. Indeed, I could believe that "MI is about to be attacked by a bear", without believing that "I am about to be attacked by a bear". This of course, will have an enormous impact on the way in which I take action on the face of the belief I happen to hold. Only if I can *anchor* the proposition "MI is about to be attacked by a bear" to *me*, am I in the position to take the action that is needed in order to avoid being attacked by a bear (see also Bermúdez 2017). That is, only if I can "translate" the non-indexical proposition in its indexical, self-referential form, am I in the position to properly act in my own interest. If, at the psychological level, the content satisfactorily expressed at the semantic level by the proposition "MI is about to be attacked by a bear" is not represented indexically, i.e. as referring to the very subject entertaining that thought/proposition, then it is impossible to see how *I* could ever take action in order to avoid being attacked by a bear.

In light of these last considerations it is possible to raise some reasonable concern about C&D's argumentative strategy of supposing the existence of a subject "Third", who does everything "third-personally". If my critique is on the right path, one might have more than one reasonable worry about how an individual like Third could ever do anything. Indeed, if Third does anything "third-personally", then how could he or she be ever in the position to engage with the world without entertaining any kind of self-referential thoughts?

That is probably why C&D specify that Third does anything third-personally "as far as this is possible". C&D do not tell us much about how far we can go in representing Third's third-personal way of navigating the world. But we can reasonably speculate about it. There is a real-life case we can turn to that could resemble in some relevant sense to the way Third would experience the world, if Third happened to be real. In the clinical literature, we find indeed the case of Ian Waterman, a British nineteen-year-old who lost most of his sensory abilities and proprioceptive capacities due to a terrible virus that infected him. Here is how neuroscientist Gregory Hickok describes his case:

> Ian was in the hospital fighting a severe virus when one morning he woke up and couldn't feel or move his body. Doctors had no idea what was going on, what his prognosis was, or how to treat him. He recovered from the virus but his body remains lost to him. In a BBC documentary called "The Man Who Lost His Body," Ian described his tragic experience poetically:
>
> "Turned every 2 hours like a joint of meat. Basted with lotions. Unmoving like a statue. Mind filled with emotion. Limbs dead to the touch, movement impossible. Lying on a bed, eyes fixed on a flaking ceiling, wishing those flakes would turn to cracks and the ceiling fall to take me from this misery. What use an active brain without mobility?"

> But Ian didn't lose mobility completely. In fact, his motor system is virtually unaffected. What is lost is his ability to feel, a condition known as large-fiber sensory neuropathy. Ian's case is extraordinarily severe. He lost all sensation from touch as well as proprioception, sensory signals from receptors deep within the muscles and joints that signal body position. (Hickok 2014: 301 -302)

What makes Ian's story even more astonishing is that, thanks to his heroically stubborn efforts, he eventually regained his ability to move and navigate the world with his body. Since his motor system was intact, he was in the position to regain control over his motor function. However, Ian had to reclaim control over his body through a different path from that usually taken by neurotypical humans: instead of relying on the usual somatosensory feedbacks, he had to rely exclusively on *visual* feedbacks.

As Hickok notices, we rarely have to rely on visual feedback in order to perform bodily movements, i.e. we do not need to check where our legs are to walk properly, simply because we have an immediate, *felt* access to our own body. This way of monitoring and controlling one's own body and actions was precluded to Ian. Therefore, in order to bring about bodily actions, he constantly had to rely on a perceptual, specifically visual, anchor to his body. This made even rather simple actions cognitively effortful for Ian:

> All [Ian]'s useful movements require constant visual control and mental concentration. In the dark he is unable to move and such is his requirement to concentrate on movement that he cannot daydream whilst walking. He finds now that walking on level terrain requires about half the amount of concentration that it did. Walking on uneven ground still demands full concentration. […]
>
> His mental concentration is finite. If in a chair he can pick up an egg without cracking the shell. But he would not be able to walk and hold the egg; for then his concentration would be on walking and his grip on the egg [would] become too firm to avoid breaking it. He drives a car with hand controls and finds it very relaxing. Here he has his hands frozen around the controls which remain in his peripheral vision, though his concentration is on the road […] All movement tasks appear to require continuous visual feedback, except for freezing in postures which are then held with frequent visual checks. (303-304)

Let us now see how this clinical case can shed some light to the arguments provided by C&D.

It seems to me that one reasonable way to interpret Ian's case is to claim that he lost his first-personal access to his own body. Ian's way of bringing about actions is in this sense third-personal: he knows that *that* the body he sees is *his* body, but at the same time the first-personal, proprioceptive access to his own body is irremediably lost. This has tremendous consequences on the ability to act in the world as efficiently as other neurotypical humans do.

Instead of immediately feeling that e.g. a potentially painful stimulus such as a burning stimulus is coming *from* his left hand so that the subsequent retractive movement can be almost mindlessly performed, an individual like Ian would have to work it out by inferring – via visual feedbacks – that a potentially dangerous stimulus *at* location l is likely to damage the body *at* location l_1, a body which has to be inferentially recognized as one's own. If, at the pure semantic level, thoughts such as "A painful stimulus at location l is likely to hurt Ian's left hand at location l_1" and "I feel a painful stimulus from my left hand" can be identical, at the psychological level, they require different cognitive processes, have different levels of efficiency and might bring about different outcomes.

Now, the case reported by Hickok can be read as a somatosensory version of Third who – as we should remember by now – does everything third-personally. Even limiting Third's third-personal access to world to the mere somatosensory dimension, is more than enough to appreciate the crucial role of the first-person mode of presentation at the level of a subject's conscious representation of the world. Let us try to imagine if we were to extend Third's third-personal access to every other domain of cognition: from deliberation to empathy. We can legitimately ask, how would Third orientate in a world without that self-referential background that (implicitly or explicitly) seem to continuously frame most of our mental life and our agency in the world? An answer is easily given: most (if not all) of Third's cognitive processes would be impaired.

7.3. Taking Stock: Perspectives Matter

The expected upshot of the above sections is twofold. On the one hand, I hope to have provided good reasons to consider the deployment of perspectives useful to convey some fundamental insights about the workings of empathy. On the other hand, I hope I also gave reasons to prove the importance of the first-person perspective in the context of our everyday cognition. In light of this, we can safely conclude that, differently from what C&D aim to show, talks of first-person perspectives are crucial to a proper framing of some of empathy's most fundamental features.

8. A Few More Words on the First-Person Perspective

In this final section, I would like to say a few more words on the first-person perspective. Most of what I will say will not sound completely new, as it has been stressed here and there in the previous sections. Still, it can be taken as a brief recap of some important aspects of an issue that is of the utmost importance for the purposes of the present work.

As we saw, there are many ways in which we can spell out first-perspectives and even those who are sceptics about their philosophical utility are willing to admit at least some talks of perspective as

legitimate. C&D were mostly interested in debunking the philosophical tendency to ascribe to the first-person perspective an irreplaceable role in the context of human agency and cognition. As we saw, their extreme criticism did not allow them to appreciate the psychological importance of first-person perspectives in the context of everyday cognition. Busy as they are in making a case for the inessential role of first-person perspectives, they fail to appreciate how pervasive and crucial they are in the general functioning of our minds. One might even agree that first-person perspectives do not play an irreplaceable role, i.e. that it will always be possible *in principle* to successfully explain every instance of human cognition and agency by deploying a perspective-free conceptual toolkit,[15] and still appreciate the role that they *practically* play in most of our cognitive processes.

Sara Aronwitz and Tania Lombrozo (Aronowitz and Lombrozo 2020) provide us with an enlightening example that allows us to appreciate the importance of the first-person perspective in our everyday cognitive processes, specifically in an inferential task concerning social cognition.

Take the present formulation:

> *Abstract.* A friend acts on a false belief that causes her to publicly violate a social expectation or norm.

Then ask yourself: how does she feel?

Now take this other formulation:

> *Experiential*. A friend attended a beach wedding in a bathing suit, only to discover the other attendees in formal attire.

And now ask yourself the same question.

I am fully confident that the last formulation is much better placed to provide us with a plausible evaluation of our friend's state, i.e. she is likely to feel embarrassed. Why is it so? What is about the last formulation that makes it more suitable for a quicker, quasi-automatic appreciation of our friend's state?

Aronwitz and Lombrozo claim that it is the *experiential* nature of the last formulation that plays a crucial role in our inferential process. In the first, "abstract" formulation the experiential features are subtracted by the description of the situation in question, whereas the second formulation is so structured that it preserves concrete details of the actual unfolding of the relevant event (e.g. sensory details such as the bathing suit of our friend). The experiential formulation is more likely to trigger

[15] I refused this claim in the previous section, but I am now merely granting it for the sake of the argument.

the empathic tendency to put ourselves in our friend's shoes, i.e. to imagine how we would feel, were we to undergo a similar experience. On the contrary, the abstract formulation does not immediately trigger any empathic response but merely requires us to produce an inference on the basis of law-like, causal links among rather abstracts variables (e.g. false beliefs, public violations of social norms, etc.).

We can translate the considerations above in our perspectival vocabulary: the abstract formulation invites us to adopt a third-person perspective, whereas the experiential formulation invites to adopt a first-person perspective on the event in question. The first-person and the third-person take *on the same event* differ in bringing about inferential processes. Specifically, the first-person take allows us to effortlessly assess how our friend would feel.

The example we focused on so far is particularly interesting because it does not even require us to go so far to claim that the first-person perspective plays a role that cannot *in principle* be taken over by any third-personal re-assessment of it. Indeed, I am willing to grant that in principle it is possible to work out our friend's state by relying on the abstract, third-personal formulation. Still, we see a difference in the two ways of representing the same event and we should be fully aware of it in order to provide a thorough description of a process like empathy. At least, we need to see how, *in practice*, the third-person and the first-person perspectives are likely to produce different outcomes or at least bear different cognitive costs, even though, *in principle*, it could still be possible to trade one for the other, *salva veritate*. This alone would already provide a solid basis to resist C&D's scepticism.

However, in the present work I am going to move a step further than that and claim that our first-person perspectives can grant us access to certain features that cannot be disclosed via a purely third-personal perspective. First-person takes on certain events do not only feel more natural in certain contexts (as we saw with Aronwitz and Lombrozo's example), but constitute the royal avenue to access certain facts about the world we inhabit, such as experiential features of the world – e.g. what it feels like to undergo certain kinds of experiences – and certain kinds of values – e.g. how do coconuts taste like? What does it feel like to see red?

In a recent paper, Katalin Balog (Balog 2020) made a distinction between objectivity and subjectivity that largely maps onto the first-person/third-person distinction I have been drawing so far:

> I understand subjectivity and objectivity in concepts, thoughts and mental processes in terms of their connection to experience. A concept, thought or mental process is subjective to the degree it is connected to experience, whereas it is objective to the degree it is abstracted from it. (Balog 2020: 1)

Subjective thoughts, states and processes or, as I prefer putting it, first-perspectives are intimately connected to experience, i.e. they preserve traces of the way in which we apprehend the world by means of subjectively experiencing it. Objective thoughts, states, and processes or, as I like to put it, third-person perspectives, instead, present their contents in a way that in principle could be available to subjects that experienced the world in ways that are radically different from ours.

Third-personal descriptions of experiences could in principle be fully understood by creatures whose experiences present a radically different phenomenological "texture" or by individuals that were not first-personally acquainted with the kinds of experiences in question.

Of course, one can translate a first-personal representation into a third-personal one and understand the former through the latter. Still, something would be lost. A subject who has access only third-personally to certain kinds of experiences will lose fundamental features of them. Most importantly, for our purposes, such individuals would not have empathy as a strategy to understand people undergoing such experiences.

When empathy is an unavailable strategy, then individuals are incapable of appreciating crucial properties and facets of experiences. These are aspects that do play an irredeemable role in making us the creatures we are.

Onwards

Assessing these basic facts about empathy, experiences and first-person perspectives constitutes the main basis onto which I am going to build the rest of the present work.

Indeed, in what follows, I am going to deep dive into many of these aspects. Specifically, I am going to explore more in detail what kind of understanding empathy is capable of delivering. Also, I will explore what kinds of epistemic accomplishment it is capable of and, also, to what kinds of shortcomings it falls prey.

In other words, I am going to explore what empathy can and cannot do.

Chapter II

WHAT EMPATHY CAN DO (PART I)

Understanding Reasons (and Beyond)

1. Introduction

In this first exploration of empathy's epistemic possibilities, I will delve into the domain of agency, more specifically the link between agency, reason and, of course, mutual understanding.

Reason is a manifestly complex topic, it has its roots in the most respectable philosophical tradition and has been widely explored by psychologists as well.[16] The present chapter will obviously not cover the entire realm of reason in a heroic attempt to enlighten a hopelessly multifarious topic. I will rather focus on a limited set of issues linked to the topic of reason, specifically on those issues that make it particularly relevant for the illustrations of empathy's specific contribution to the problem of human mutual understanding.

As noted by Christine Korsgaard (Korsgaard 2008) people have at least three things in mind when they talk about reason:

i) Reason as a form of reflective, active thinking which is usually to be contrasted to other more receptive and passive forms of thinking such as perception and, sometimes, emotion. According to this view, perception and emotion constitute more automatic and reflex-like ways of processing the surrounding environment whereas reason is a "self-conscious, self-directing activity through which we deliberately give shape to the inputs of receptivity. This happens both in the case of theoretical reasoning, when we are constructing a scientific account of the world, and in the case of practical reasoning, where its characteristic manifestation is choice" (Korsgaard 2008: 207). In other words, reason can be conceived of as a form of epistemic vigilance that carefully mediates both our theoretical and practical activities.

ii) Reason as conformity to certain domain-general principles such as the rules of logic, mathematical principles or other rules of practical reasoning. In this sense, a person is

---

[16] For a contemporary landmark publication on the topic of reason see (Sperber and Mercier 2017).

called more or less "reasonable" depending on her degree of compliance with the aforementioned principles.

iii) Reason as particular considerations that count in favour of holding certain beliefs or bringing about certain kinds of actions.

Throughout the rest of this chapter, I will be focusing mainly on the third notion of reason and more specifically to its relevance for agency. Namely, reasons will be taken as those "facts" about human psychology that relevantly "speak for" certain kinds of behaviours and actions brought about by subjects.

Before moving forward in our analysis, it should be noted, though, that the distinctions I just drew with the help of Korsgaard's analysis are not to be conceived as rigid and mutually exclusive categories. Indeed, even if it is easy to see how the three enlisted notions of reason differ from one another, it is equally simple to foresee ways in which they can seamlessly interact with one another. One example could help us here. Imagine that, on the way to my office in the streets of Paris I bump into a fanciful *boulangerie'*s window that is displaying some freshly baked *pains au chocolat*. After some moments of ecstatic contemplation, I firmly move away from it and keep walking my way to the office. Why did I act the way I did? Certainly, there was a reason that made me behave in such a way, i.e. a reason in the third sense (iii) moved me away from the *boulangerie*. We can now ask more specifically what *kind* of reason did I have for not buying the *pain au chocolat*? First, I had to resist my temptation to rush into the store and buy a dozen of them, i.e. I had to deploy reason as described in (i) to refrain my automatic tendency to eat the *pain au chocolat* and exert a form of more thoughtful vigilance on my instincts. What kind of principle took the place of my instinct? I realised that I just had breakfast and that another injection of sugar and saturated fat in my body would definitely be far from an ideal standard of healthy habits to which I tend to conform. Here, is reason as described in (ii) that is playing a decisive role in the way I act. Specifically, it is my conformity to practical principles of healthy eating that are playing a decisive role in providing me with a suitable reason to move away from the bakery.

As we just saw, the three categories of reason we just individuated can be porous with respect to one another, e.g. what defines a reason for action (iii) can be to notions of reason as described in (i) or (ii). At the same time, we have to be fully aware that reasons for actions can be found in facts and psychological states that are not reducible to the instance of reasons as described in (i) and (ii).

As we will see, indeed, one of the main focus of this chapter will consist in illustrating how emotions, which – as we saw – cannot be plainly included in notions (i) and (ii) of reason, play a substantive

role in shaping our agency and how an empathic understanding of people's reason for action has, in certain relevant contexts, to pass through a proper understanding of the emotional states people are in when behaving as they do.

But let us proceed in order and try to clear up our minds more in detail about how to conceive of reasons for action.

2. Reason = Belief + Desire?

Take some very common behaviours: our spouse opening the fridge on a hot summer day to grab a bottle of water. Our friend Jane taking her car keys out of her purse as we are approaching her car. These are familiar instances of behaviours we can effortlessly make sense of in the blink of an eye. More specifically, we can see these behaviours as being occasioned by certain mental states held by the subjects who bring them about and that provide suitable reasons for the displayed behaviours.

An authoritative philosophical tradition has been unusually unanimous in identifying said mental states with beliefs and desires held by agents. Namely, it is because our spouse *wants* to drink some water to quench her thirst and *believes* that there is a bottle of water in the fridge that she behaves as she does, and it is because our friend *wants* to open the car and *believes* that the proper means to open it that she picks it from her purse as we approach her car. Being able to attribute the relevant belief and desire couple to agents is what allows subjects confronted with their behaviours to make perfectly sense of them in terms of *reasons*: when provided with the relevant desire and the suitable instrumental belief, we are – in principle – in the position to understand, explain, and predict other agents behaviours.

The philosophical debate has been rich in debates about the link between rationality and practices of mutual understanding (see e.g. Davidson 1963; Dennett 1987; Heal 2003; Millar 2004). The intersection between these two topics, though, is too broad to be covered in the span of the present work.

I will rather concentrate – as it is reasonable to expect – on a smaller subclass of mindreading, namely empathy, in its relationship to the understanding of reasons agents have for their behaviours.

In the contemporary debate, one of the most influential and sophisticated attempts to explore the link between empathy and reason can be found in the work of Karsten Stueber (e.g. Stueber 2006). Interestingly enough, the account of empathy he relies on when discussing the role of reason in empathic processes is relevantly similar to the one that I presented in the previous chapter.

He describes empathetic perspective-taking processes – which he dubs as "re-enactive empathy" – in this way:[17]

> [Re-enactive empathy is a process] that requires us to adopt an alternative to the reality of our perspective by transposing ourselves imaginatively into the shoes of another person. (Stueber 2016: 327)

This kind of empathy is called re-enactive because the subject performing it has to almost literally "re-enact" in her mind the relevant mental states held by the person she is intending to empathize with. I can re-enact a person's belief that there is a bottle in the fridge and their desire to drink some water by imaginatively holding such mental states. As a result of it, Stuber says, I am provided with an empathic understanding of the *reason* responsible for my that person's behaviour.

I am going to call this model of empathy the *agential conception of empathy* or, simply, *agential empathy*. This model focuses on empathy as a mental process capable of providing insight into the reasons agents have to behave the way they do in a given situation. Interestingly enough, as we just saw, the agential conception of empathy put forward by Stueber is largely aligned with the aforementioned philosophical traditions that identify reasons as being constituted by a belief and desire couple. This model is usually known in the specialized literature as the "Humean model" of action explanation.

As I will try to show in what follows, I believe that the Humean model is incomplete. Indeed, representing reasons in terms of beliefs and desires leaves out another crucial psychological factor that plays a pivotal and irreplaceable role in shaping human agency, namely emotions. Stueber, by largely committing himself to a reductive model of reason, leaves the category of emotion largely unexplored. Not much is to be found in his work concerning *how* emotions can motivate us to act and, therefore, provide us with compelling, motivating reasons to bring about certain behaviours.

My goal will mainly consist in doing justice to the role of emotions as full-right reason-giving states: if we want to understand empathy as capable of providing us with an understanding of reasons for acting in a certain way, then we have to be willing to include emotions among those mental states that empathy is capable to re-enact, i.e. to imaginatively represent from a first-person perspective.

---

[17] Stueber usually distinguishes between two kinds of empathy, basic empathy and re-enactive empathy. Basic empathy is conceived of as the empathizer's capability of understanding other individuals as minded creatures, i.e. creatures with mental states such as e.g. emotions. This kind of empathy is taken to be mostly accomplished by built-in, sub-personal mechanisms such as mirror neurons. We will leave basic empathy out of our focus here. Indeed, it is re-enactive empathy that is relevant for our discussion of reasons in the context of the empathic understanding of agency. For a similar distinction see also Goldman's account of low-level and high-level simulation (Goldman 2006).

As we will see, though, accounting for the role of the emotion as reason-giving states will allow me to highlight an aspect of empathy that has been almost completely ignored by Stueber and that is still somewhat neglected in the specialized literature on empathy. Indeed, if we want to properly comprehend empathy as capable of providing an understanding of agency through the lens of reasons, then we need to stress its capacity to convey insight into the experiential dimension of our mental states. As I will try to show in the coming sections, emotions specific motivational force is to be found in their phenomenology.

Acknowledging empathy's capability of providing insight into the experiential dimension of our mental life, i.e. what we introduced in the previous chapter as *phenomenal insight*, will allow us not only to carve the required space for emotions within the agential conception of empathy presented by Stueber, but also to fully envision for empathy an epistemic function that Stueber has largely neglected.

3. Re-enactive empathy and understanding reasons

As already announced, Stueber inscribes his account within a "rationality assumption" according to which we, as agents, are able to understand each other because we see ourselves as acting out of reasons, i.e. as rational agents (e.g. Stueber 2011). This basically means that agents, rather than being inert entities tossed around the world by external influences, are usually in the position to say something *in favour of their actions* (Stueber 2017). Namely, actions brought about by a rational agent are placed within a reason-giving explanatory framework that makes them intelligible to themselves and, consequently to other rational agents like them.

As written by Stueber:

> Rational agents are not merely creatures who act because something is happening inside them. Rather, they are able to take a reflective stance towards their own agency and to take ownership of their actions in terms of their reasons for acting. (Stueber 2012: 59)

The amenability to reason of an agent's behaviour is what makes room for the existence of *intentional* action and what ultimately substantiates our everyday practices of mutual understanding. Indeed, it is in virtue of the reason-responsiveness of our actions that we are able to make sense of each's behaviours.

But how are these reasons to be conceived of? As already announced in the previous section, at the psychological level, they have been usually taken to be instantiated by a belief and desire couple held by the subject performing the action.

An intentional action is motivated by a desire and a relevant instrumental belief that a certain action (namely, the action that is actually brought about) is the best way to fulfil the desire. Stueber's account makes re-enactive empathy central in allowing mindreaders to grasp the reasons that govern another subejct's actions.[18] Interestingly enough for our present purposes, Stueber's account of empathy is largely aligned with the standard view of reasons as constituted by beliefs and desires. Indeed, empathy, as a form of imaginative perspective-taking, is what allows a subjects to "try on" the relevant mental states held by a person in a given situation – namely, beliefs and desires – so that they can get acquainted with the *reasons* that lead a certain individual to behave the way they did. By imagining holding the relevant mental states that are taken to be held by agents in given circumstances, empathizers can readily grasp the reason-giving engine of their behaviour.

Coming back to our previous example, if I imagine holding the desire to drink some water and the belief that there is some water in the fridge, I am immediately provided with an understanding of the "first-personal push" that those states can exert on my subsequent behaviour, i.e. the *reason* for reaching the fridge and grabbing some water becomes fully intelligible to me.

Now, as I will try to show in what follows, there are some good reasons to be dissatisfied with a theoretical move that conceives of the realm of reasons exclusively in terms of beliefs and desires. Indeed, emotions do play a fundamental role in providing subjects with *reasons* to bring about certain actions. Therefore, if we are to conceive of empathy as capable of providing us with an understanding of reasons, we need to include emotions within the epistemic scope of empathy. This, as it will become clearer in what follows, we need to ascribe to empathy not only the capacity to re-enact the mere *propositional* contents of certain mental states but also their relevant phenomenal properties. This turns out to be particularly relevant in the case of emotions: emotional states, indeed, compel us to act in a certain in virtue of the way in which they *feel*, i.e. the way in which they are experienced from a first-person perspective.

[18] From now on, I will deploy the term "empathy" to refer to Stueber's re-enactive empathy which, as far as the present discussion is concerned is relevantly similar to the account I have been presenting in the previous chapter. This, of course, does not mean that I fully embrace the account provided by Stueber. By focusing on the similarity, I am in the position to critically engage with the aspects of his theory that are particularly significant for my argumentative purposes. But this should not induce the reader to believe that I buy the entire empathy-package provided by Stuber. For instance, his manifest commitment to simulation theory and to the idea that empathy represents our default and most pervasive mindreading strategy, are claims which I tend to resist. However, since the scope of my work does not consist in a critical exploration of Stuber's account of empathy, I can ignore these concerns..

4. Arational actions

A convincing case against the belief-desire model of reason can be made by looking at the case of what Rosalind Hursthouse dubbed as "arational actions" (Hursthouse 1991). These actions are particularly interesting because they cannot be accommodated by resorting to any suitable belief and desire held by the agent who performs them. Instances of arational actions include – but are not limited to – some of the following paradigmatic cases:

> (i) Rumpling someone's hair (out of love)
> (ii) Gouging holes in someone's picture (out of hatred)
> (iii) Kicking a door (out of anger)
> (iv) Jumping up and down (out of joy)
> (v) Rolling around in one's dead wife's clothes (out of grief)
> (vi) Covering one's face in the dark (out of shame)
> (Scarantino and Nielsen 2015: 2977)

They are defined as *arational* because they cannot be explained by means of reasons, if reasons are to be understood as being solely constituted by a belief-desire couple. In the specialized literature, this peculiar category of actions is also known under the terms "expressive actions" (Döring 2003, 2007) or "genuine expressions" **(**Goldie 2000**).** [1]

As noted by Hurthouse, we can make these actions fully intelligible only if we interpret them to be motivated by an emotion. As she puts it: "in the grip of the relevant emotion, the agent just felt like doing them [viz. the aforementioned emotional actions]" (Hurthouse 1991: 61).

Hurthouse individuated three necessary and sufficient conditions for arational emotional actions. We can claim that an action brought about by an agent was arational we can say about it:

> (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 he 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.

We can unpack the meaning of the intricated prose of Hurthouse by saying that arational actions are (i) intentional, (ii) not amenable to an interpretation in terms of beliefs and desires, (iii) done in the grip of an emotion.

Before firmly rejecting the belief-desire explanation, Hurthouse reviews two possible responses that might be articulated by a tenacious defender of Humean model.

Let us focus on one of her most discussed example provided by Hurthouse: "Jane (…) in a wave of hatred jealousy for Joan, tears at Joan's photo with her nails, and gouges holes in the eyes" (Hurtouse 1991: 59). Humeans could try to explain such behavior in three possible ways:

(i) They ascribe to Jane the desire to truly harm Joan and the belief that by scratching a photo of her would enable her to achieve this goal.

We should reject such a solution because we would be forced to postulate the existence of an absurd instrumental belief, i.e. the belief that scratching a photo of Joan would enable Jane to harm her. Even if it is fully plausible that some agents could believe in some mystic Voodoo properties of the picture, it is also equally true that an agent who does not hold any Voodoo belief could still bring about such an action. Humean should be able to account for the occurrence of such an action in the absence of such an unusual belief.

(ii) Jane wanted to express her hate against Joan and believed that scratching Joan's photo was the right way to express her emotion.

This explanation is clearly untenable because we the belief we ascribe to Jane implies that the action can be submitted to an evaluation in terms of correction. Since we cannot say that an agent *correctly* (or *wrongly*) expresses her emotion, we cannot apply such a belief in this context. It would simply be preposterous to claim that people in Jane's situation are actually trying to correctly express their emotional state. People do not usually perform arational actions in the attempt to correctly express their emotions.

(iii) Jane did this because she desired pleasure and believed that such an action would enable her to do so.

This final explanation is also unsatisfying. Indeed, we do not see how gouging holes in a picture would necessarily bring her pleasure and relief. It is surely possible for certain instances of this action to bring pleasure. But at the same time it sounds perfectly plausible to imagine Jane behaving the way she did, even without ascribing to her the desire for pleasure and the (somewhat bizarre) belief that gouging holes in a picture constitutes a pleasurable experience.

It seems that, at this point, the only tenable solution would consist in describing Jane's action as brought about in the grip of the emotion of hatred. In other words, the *reason* for her action is to be

found in the emotion she is undergoing rather than in any implausible belief and desire she might happen to implicitly or explicitly hold.

4.2. Goldie's Humean tenacity

A further, tenacious attempt to provide a Humean explanation for arational action has been provided by Peter Goldie (2000). Goldie indeed believes that, in cases of arational actions, a Humean explanation is still available, it simply needs to be more sophisticated than usual. In order to make sense of Jane's action we need to ascribe to her also the desire to harm Joan. Goldie's interesting move consists in conceiving of such a desire as taking the form of a wish, i.e. a desire whose satisfaction involves an act of the imagination:

> In the sense in which I will use the term, when I wish for something, I desire that thing, and also I imagine, or I am disposed to imagine, the desire to be satisfied. In this sense, not all idle desires need involve wishes, as I might not, for example, be disposed to imagine myself being taller; and not all wishes need involve idle desires, as I might imagine doing something which I want to do, and which I could, in fact, actually do—hitting someone, for example. (Goldie 2000: 129)

This move allows Goldie to attribute to Jane the wish to harm Joan which takes the form of a desire to actually harm her plus the imaginative act that tearing at her photo will enable her to do so. Jane can only wish to do so because certain "civilizing restraints" make the hypothesis of actually blinding Joan not an available option – at least not available for most people. Therefore, Jane would attempt to satisfy the socially unacceptable desire to blind Joan symbolically by gauging holes in Joan's picture. In this sense then, Jane's behavior is the *sublimated* version of the action of actually harming Joan:

> The civilizing restraints on what an angry, hating person can do—ethical restraints perhaps, or knowledge of the force of the law—are just what makes Jane perfectly aware that she ought not to do bodily harm to Joan, thus leading her, on this occasion, to resort to an expressive action. The symbolic nature of the expression takes place as it does partly because the literal action, as it were, is not a realistic option. (Goldie 2000: 130)

There are several reasons to be dissatisfied with Goldie's Humean move. First, it is not always possible to find a symbolic level of explanation for certain instances of arational actions. Indeed, it is not clear at all what would be the symbolically fulfilled wish behind an action such as "kicking the table out of anger" (Scarantino & Nielsen 2015). Second, as noted by Adam Kovach and Craig Lancey (Kovach and Lancey 2005) Goldie's solution is explanatorily deficient: it plainly fails to provide justification for the supposedly motivating efficacy of wishes. Indeed, there is plenty of wishes we can think of that blatantly fail to motivate us: I wish I could concentrate way more than I

usually manage to, or I wish I could run a full marathon. These are authentic wishes that are utterly unsuccessful to motivate any instance of behaviour towards their fulfillment: I am still passively prey of meaningless distractions and I haven't run for months. Third, Goldie's solution seems to require the co-occurrence of two relevant desires (the desire to tear at the photo and the desire to harm Joan) and a belief in order to account for the occurrence of Joan's action. So, one might finally have some reasonable reservations about Goldie's explanation actually fitting with the Humean prescription according to which reasons for actions are to be found in a single desire associated with a single instrumental belief.

4.3. Stueber's Humeanism

Before moving forward in my critical evaluation of Humeanism within the context of empathic perspective-taking and presenting a suitable solution to the issues I just raised, it might be useful to come back to Stuber's re-enactive empathy to better assess his position with respect to the belief-desire model.

I believe that it would certainly be uncharitable to attribute to Stueber an overtly reductionist view where reasons are *nothing but* the result of a relevant belief-desire couple. Indeed, while it appears clear that his arguments are primarily calibrated on cases that plainly fit with the Humean take on reasons, it is true that in some passages of his work he acknowledges the possibility to empathically re-enact mental states other than beliefs and desires, such as emotions:

> Goldman presents us with the case of somebody who just missed a train (it left one minute before he reached the plat- form) compared with somebody who misses a train by two hours. We intuitively understand that the person who just missed the train is more annoyed. […] What we do understand—using the method of simulation—I would like to suggest, are the aspects of the situation that constitute our reasons for responding emotionally in a certain way or, to express it in a more neutral manner, the aspects of the situation that would allow such a response to be appropriate. We grasp that the person who just missed the train has more reason to be annoyed than the person whose train left over an hour ago, because if the latter, say, would have avoided wasting time talking to the cab driver, or would have run or driven a little bit faster, he probably would have made it in time. (Stueber 2006: 161)

In this passage, Stueber grants the possibility of empathically matching the emotional state of a person who just missed the train by imaginatively putting ourselves in her situation. Empathy allows us to first-personally engage with the relevant features of a certain event that *warrant* or *provide reasons for* an emotional response of a certain kind. Imagining myself occupying the perspective of the person who just missed the train, allows me to immediately understand what aspects of his situation are

likely to trigger an anger response, that is, what features of the target's perspective provide *reasons* for her anger:

> As Aristotle correctly and repeatedly pointed out, the ability to respond in a specific situation in such a manner presupposes that one is emotionally attuned to the world in the right manner. [...] We ordinarily judge an emotional response to be appropriate or inappropriate; we recognize, as Aristotle put it, that it is appropriate to feel angry at the right time, the right place, and toward the right person. We indeed take it to be a good reason for you to be angry if somebody intentionally steps on your foot to hurt you. But we would not find this reaction appropriate in the case where somebody steps on your foot by accident. (2006: 160–161)

In this other passage, Stueber argues that it is possible to assess emotional states in reason-giving terms or, if we prefer, according to some (broadly construed) principles of rationality, as far we can see their occurrence as instantiated by some states of affairs that – at least partially – motivate them.

I do not have anything to oppose to this view, but I still believe that there is a whole story to be told about how emotional states can shape our behavior and how empathizers are able to re-enact these mental states to properly understand their efficacy in other people's actions and mental life more generally. Indeed, as it is certainly true that emotional states can be grounded in reasons, it is also true that they can operate as main reason-providers in our agency.

I believe that Stueber's account remains completely silent on this latter aspect of emotions. Therefore, his model remains fundamentally incomplete. It can account for cases that nicely fit the Humean framework. When it comes to more controversial cases, as we saw with the case of arational actions, we are left in need of further explanation. So, if we want to include emotions within an account of empathy as capable of providing understanding for subjects' reasons for actions, a few considerations are still to be made. I will turn to them in the next sections.

5. Accounting for the experiential dimension of emotions

The case of arational emotional actions highlighted a blind spot in Stueber's account of re-enactive empathy. As we saw, arational actions are not amenable to a Humean account of reason. We are then forced to look somewhere else if we want to grant to empathy the possibility of delivering understanding on the reason-giving features of emotional states. We can express this point more schematically:

i) If we take reasons as reason-giving mental states in their full rights, irreducible to the belief-desire model;

ii) And, if we take empathy as capable of providing us with an understanding of reasons that "speak for" a subject's action in a given situation;

iii) Then, we still need to specify *how* empathy is capable of providing insight into the motivational, reason-giving features of emotional states.

I will claim that empathy is capable of grasping the motivational component of emotional states by providing us with an insight into their phenomenological dimension.

In order to adequately illustrate this point, I shall delve into some relevant debates in emotion theory that will allow me to target the specific role of emotions in re-enactive empathy and, more specifically, to account for the role played by their phenomenological component. These considerations will turn out extremely useful to illustrate how we can empathize with arational emotional action. I will then extend these considerations to the case of emotional actions in general.

5.1. Affective phenomenology and motivation

A promising way to account for the case of arational actions has been offered by Sabine Döring (2003, 2007). She is on board with us in holding that we cannot adequately account for arational actions purely in terms of beliefs and desires and that we necessarily have to resort to the emotional state of agents. Her account is particularly interesting because it relies on the phenomenological component of emotional states in order to account for their motivational force. Engaging with her account will enable me to highlight some interesting issues related to empathy's reason-giving explanatory power.

It should be noted that Döring counts among the proponents of a perceptual theory of emotion according to which emotions can be regarded as a specific kind of perception.[19] Both perceptions and emotions, indeed, figure as intentional states that are irreducible to other mental states due to their specific phenomenology. Importantly, as perceptions give us access to fundamental properties of the world such as spatial and color properties, emotions are states that give us access to certain evaluative properties at the level of their contents. The way in which emotions present specific evaluative properties is through their subjective character, i.e. their phenomenal dimension. Emotions non-inferentially reveal to us certain aspects of the world that need not be primitively based on nor constituted by the presence of other evaluative judgments.

[19] Other scholars in the field have defended a perceptualist account of emotions, (e.g. Goldie 2000; Johnston 2001; Prinz 2004).

The so-called recalcitrant emotions provide us with an apt example of how this might be the case: suppose I am so afraid of spiders that simply looking at them is enough to instill a profound emotion of fear in me that urges me to move away from the stimulus. No matter how innocuous the spider is – it might even be a spider-shaped plastic toy – and I can be fully aware of its harmlessness – i.e. I might *believe* or *judge* that the spider is innocuous –, I will still experience fear and I will still move away from it.[20]

Let us come back to the case of Jane and Joan. In light of what we said so far, we can explain Jane's action of tearing tat Joan's picture out of hatred is grounded in Jane's *felt* representation of Joan as being an obnoxious person. The motivational force to scratch Joan's picture and gauge holes in her eyes is rooted in the emotion's *affect*, i.e. in the felt dimension of the emotional state. Reference to the emotion's affect is what allows us to provide a reason for Jane's behavior. In Döring's words:

> Explaining an action by specifying how the emotion's affect relates to its representational content in causing the action. It is the emotion's affect which gives it motivational force [..]. Emotions are capable of motivating because their representational content is at the same time felt […]. (Döring 2033: 223-224)

Coming back to our spider example, it is the very phenomenology of fear that compels me to flee whenever I encounter a spider: I represent the spider as something *to be afraid of* and that representational content is to be found in the what-it-is-like character of the emotion of fear. As a result, I tend to act according to the evaluative aspect non-inferentially unveiled to me by the emotion's affect, i.e. I move away from the fear-inducing stimulus.

What is crucial, with respect to the Humean theory of motivation, is that emotion's motivational push can be independent of any desires the agent might have. In Jane's case, her emotion of hatred should motivate her to harm Joan directly but, given the social and moral constraints that make it inconvenient to harm other people, Jane limits herself to gauging holes in Joan's picture. That is, the emotion fails to provide an end to her action. In this sense, the means-end logic that usually applies to desires (in tandem with beliefs) in bringing about actions collapses. For that reason, Döring concludes, the emotional state cannot be reduced to any more primitive desires Jane might have. Since we have no reason to tell emotional states that occur in the case of arational emotion apart from other, more usual, occurrences of emotional states, as if they were a *sui generis* class of emotional

[20] Such cases have been addressed from several angles in the philosophical and psychological literature. Probably, from a philosophical point of view, the cornerstone treatment of them has been offered by Tamar Gendler (Gendler 2008) who influentially referred to them as cases of "alief".

phenomena, Döring concludes that these considerations should be extended to emotions in general. Indeed, it makes no sense to introduce an undue fragmentation in the emotion category based on their amenability to a belief-desire re-interpretation. Indeed, the same emotional occurrence of, say, anger can induce us to bring about arational actions (e.g. kicking the wall, punching the door) or rational actions (e.g. fighting with someone). Therefore, we can safely conclude that emotions can, and in fact often *do*, motivate us to act independently of any belief and desire we may hold.

5.2. Back to phenomenal insight: Arational emotional actions

From what I have said so far, it seems fair to infer that in order to empathize with arational emotional actions, the empathizer has to imaginatively re-enact the agent's emotional state. More precisely, it is by imaginatively representing the phenomenal dimension of the emotion that we can empathically understand the *reason* behind an emotional action.

I introduced, in the previous chapter, the notion of phenomenal insight in order to account for empathy's capability to convey understanding about the *experienced* dimension of mental states. It is precisely this notion that will allow us to integrate Stueber's interpretation so that it can account for the problematic case of arational actions. Indeed, it is because we can become aware, through phenomenal insight, of what it feels like to undergo a certain kind of emotion that our arational actions become intelligible. Someone who has never experienced in her life the fury of anger would not be able to fully understand from a first-person perspective why Jane is doing such a purposeless action.

We can easily imagine a Spock-like character who, when confronted with Jane's behavior, dismissively categorizes it as another bizarre instance of human extravagance. This Spock-like figure, thanks to his longstanding experience with humans, could even cease to be struck by these apparently purposeless human conducts and come to formulate an empirical generalization like the following one: "humans tend to behave purposelessly when they are in the grip of emotions". This rough and ready rule of thumb would allow him to make sense of arational actions in a very minimal way, i.e. one that simply prevents him from being struck by their occurrence every time he faces one. Still, the emotion-blind Spock-like character, would lack *empathic* understanding of Jane. This strategy would simply be unavailable to him. He would fail to understand how, from a first-personal perspective, an emotional state, say of anger, can provide someone with a reason for an apparently purposeless action, e.g. gauging hole in someone's picture.

Mark Johnston insightfully illustrated this point by referring to what he calls "the authority of affects":

> By "the authority of affect" I mean not to refer to its sheer effectiveness as a source of desire or action, but rather to the fact that the presence of the affect can make the desire or action especially intelligible to the agent himself. It can make the desire or act seem apt or fitting in a way that silences any demand for justification. (Johnston 2001: 189)

The phenomenology of affective states, in our case of emotions, is what makes certain kinds of desires or actions readily intelligible in a way that silence the demand for any further justification, i.e. they disclose certain evaluative facts that make a certain kind of desire or action particularly intelligible to the agent himself and, I would add, to the empathizer as well.

Phenomenal insight, in this context, is what leads the empathizer to pay attention to the phenomenal aspects that accompany the mental states of the agent with whom she empathizes. Phenomenal insight, by providing us with a dive into the very felt dimension of agents' experiences, can substantially contribute to the understanding of an agents' reason for acting the way she does. This turns out to be particularly evident in the case of arational emotional actions.

Now we can ask: is phenomenal insight required also to fully understand the reasons behind *rational* emotional actions?

5.3. Back to phenomenal insight: Rational emotional actions

At first blush, rational emotional actions do not seem to require the full intervention of phenomenal insight to disclose the experiential component of the emotional state because they seem to be amenable to a purely Humean interpretation.

For instance, let us take the example of Sally, who runs out of the shopping mall because of a powerful earthquake. We can perfectly make sense of this behavior without having to mention the eventual state of fear she was in: the *reason* for her running away is to be found in the desire to protect herself and in her belief that the best way to accomplish this goal is by running away. I believe that this solution is unsatisfying if we grant that Sally ran away from the mall *out of fear*. Indeed, if Sally was actually experiencing fear and if fear played a pivotal role in motivating her to run away, then we are definitely leaving something out by limiting empathy's re-enactive work to the propositional states of beliefs and desires. Indeed, as we largely made clear in the previous chapter, empathy aims at representing the first-personal perspective of a subject in a given situation. If, in the empathic process, we are concocting belief and desire states while leaving the emotional component largely outside the empathic representation, then we are rather *mis*-representing our target's perspective. Stueber himself grants to empathy the capacity to faithfully reproduce the other person's mental process:

> We grasp another person's action as a rationally compelling one because we can grasp his thoughts as reasons for acting by putting ourselves in his shoes, by imagining the situation that he faces and trying to reenact his thought processes in our mind. (Stueber 2012: 28)

Now, one might reasonably wonder, we are not actually reenacting anybody's process if, instead of engaging with the target's actual mental states, we are concocting *other* mental states that do not match their real first-person perspective. Coming back to Sally, we are not re-enacting her thought process if we are leaving the emotion out of our representation! Omitting Sally's fear would not provide us with an actual insight into *how* she was motivated to act the way she did. Moreover, if we embrace the restrictive Humean view on empathy, then nothing would allow us to distinguish the reenactment of Sally's perspective where she acts out of fear from Spock-like Sally's perspective, where she acts coldly, without experiencing any emotion.

In order to make the necessity of representing the first-personal dimension of emotional states in cases of rational actions even more evident, we can take into account another example. Imagine a case where Tom and George are on hiking trial, in front of them there is a rickety bridge. George resolutely crosses the bridge without blinking an eye. Tom remains paralyzed where he is, incapable of moving one single step towards the bridge. We can now suppose that both George and Tom hold the same relevant belief-desire pair regarding the situation they are facing: they both *desire* not to be harmed by a situation that they both *believe* it to be dangerous. At the same time, George crossed the bridge, Tom retraced his steps. From a purely Humean perspective, we would be left puzzled by the asymmetry in the two friends' behavior.

Of course, the solution to this fake puzzle is easily attained if we broaden up the somewhat caustic Humean horizon on rational agency. In Tom's case, it is *fear* that led him to act so as to avoid a situation that he not only evaluated as dangerous but also as *to be avoided*. In other words, Tom perceived the situation as affectively charged, through the lens of fear, and this motivated him to act accordingly. We, as empathizers, need to re-enact, through phenomenal insight, Tom's experience of fear in front of the bridge in order to make his behavior fully intelligible – especially when considered against the background of George's behavior.

Putting the pieces of the puzzle together, in light of what I have been claiming throughout these sections, we are led to conclude that the reenacting emotions – and as I argued, especially their phenomenal components – is the only way to make it fully intelligible as a reason-giving state in its full right, i.e. as what makes us fully understand the reason behind corresponding emotional actions. This implies that Stueber's account of reenactive empathy as a royal avenue to understanding reasons

for agency needs to be integrated by acknowledging the pivotal role of emotions and phenomenal insight in understanding agency. Stueber – and Humeans in general – problematically flattens the significant phenomenological differences between those mental states, without doing justice to their specificity. In his model, recreating an emotion in our mind and attributing it to others does not seem to be much different from reenacting the corresponding beliefs and desires that could explain it. As we saw before, though, beliefs and desires on one side, and emotions on the other, do not play the same role in explaining emotional actions.

6. Some unsolved issues

Before concluding this chapter, I need to take care of some perplexities that could be raised in response to the arguments that I have presented so far. We can group them into three main issues:

(i) The problem of unconscious emotions.
(ii) Beliefs and desires as fictional constructs in service of *post-hoc* rationalization practices.
(iii) Imagining emotions: how do we do that?

I will take care of the first two issues in the next two paragraphs, whereas I will deal with the third one more extensively in the *Appendix* at the end of this chapter, as the topic of emotion and imagination deserves a longer and more articulate treatment.

6.1. What about unconscious emotions?

William James famously wrote:

> What kind of an emotion of fear would be left, if the feelings neither of quickened heart-beats nor of shallow breathing, neither of trembling lips nor of weakened limbs, neither of goose-flesh nor of visceral stirrings, were present, it is quite impossible to think. Can one fancy the state of rage and picture no ebullition of it in the chest, no flushing of the face, no dilatation of the nostrils, no clenching of the teeth, no impulse to vigorous action, but in their stead limp muscles, calm breathing, and a placid face? The present writer, for one, certainly cannot. (James 1884: 194)

We do not have to embrace James' general theory of emotions to appreciate the persuasiveness of this quote. Independently on whether or not we are willing to identify an emotion's phenomenology with the perception of bodily states that typically accompany it, we can nevertheless be persuaded by the claim that undergoing an emotional state has a distinctive phenomenology, i.e. that there is something that it feels like to undergo an emotion. However, even if it is intuitively appealing and made even more aptly persuasive by James' elegant writing, this idea has been challenged several times throughout the history of psychology: not only psychoanalysis but also certain contemporary

outputs of experimental psychology have given more credibility to the initially intuitively odd idea of unconscious emotions.

Before addressing some philosophical issues related to the problem of unconscious emotions, I first have to make explicit why their occurrence could be problematic for the account of empathic reenactment of emotions I sketched in the previous sections.

The reason is very simple: if we ascribe the motivating efficacy of emotional states to the very phenomenology of emotions, how can we account for instances of emotional states that do not have any phenomenological feature? What are we to make of unconscious emotions? Indeed, if unconscious emotions can shape the cognitive appraisal of our environment and exert their motivational efficacy in ways that are utterly independent from any phenomenally conscious feel that accompanies them, then the theoretical basis for the arguments presented above turns out to be highly unstable. An unconscious fear, one that makes me avoid a situation without me undergoing any fear-related phenomenal state, would turn out to be utterly untraceable for empathy's radars, while at the same time being prominently involved in shaping an agent's behavior.

But first of all, we have to clarify what scholars in the field mean when they talk about unconscious emotions. As recently noted by Demian Whiting (Whiting 2018), when addressing the issue of unconscious emotions, people can usually refer to two different things: either they refer to some emotions that have a dispositional profile, such as the fear of height; or they refer to episodic instances of emotions that occur below the threshold of consciousness. Dispositional emotions, such as the love for our spouse, or the hatred for an enemy, or the fear of height are enduring dispositions of our personality. Instead of having the episodic, short-lived psychological profile of many other emotional occurrences, they seem to be mental states that can be ascribed to subjects throughout a fairly extended lifespan. Suppose that I am afraid of height, then you can certainly ascribe to me the fear of height at any moment of the day and probably at any other moment of my life. But here is the trick: if it is fair to say that I am afraid of height even when I am cooking diner or completely lost in the depths of a dreamless sleep, then we have to conclude that these emotions can be unconscious: in these cases there is clearly nothing that it is like to be afraid of heights. With a punchline: there is nothing that feels like to love my partner when I am busy filling the tax forms.

I believe there are excellent reasons to be dissatisfied with this argument. First of all, as noticed by Whiting (2018), this view commits us to an "unhappy metaphysics": it would require us to claim that one entity, i.e. emotion, has two different natures, i.e. it is both an episodic mental occurrence and an enduring disposition. Therefore, either we embrace the untenable position that denies to emotion the

status of episodic mental states, or we deny that they can be dispositional. The latter is by far a preferable option. Indeed, we can argue, together with Julien Deonna and Fabrice Teroni (Deonna and Teroni 2012), that the fear of heights or the love for one's partner are *dispositions* to undergo certain kinds of emotions and not emotions themselves. If we embrace this position we do not have to commit to a problematic metaphysics and we can still hold that emotions are *felt* mental states, i.e. that they are conscious occurrences with a distinctive phenomenological profile. In fact, my fear of height is nothing but a disposition to undergo a felt episodic emotional occurrence of fear whenever I am confronted with heights.

Another strategy to avoid the threat is offered by Whiting (2018: 316-317). Instead of depriving phenomena such as the fear of height and the love for our partner from the status of emotion and classifying them as mere emotional dispositions, he is willing to claim that these states are indeed emotions in their full rights. He then asks us to distinguish between the *emotion* and the *having of the emotion* and concludes that only the latter is a disposition and not the emotion itself. As an example, to strengthen his point, he refers to the case of birds who *have* alarm calls. In such a case, it is correct to say that birds do have alarm calls when they are sleeping or mating but this does not commit us to infer that alarm calls are dispositions. Similar considerations can be made for emotions such as the fear of heights or the romantic love for someone. These can be interpreted, indeed, as being fully-fledged emotions, rather than dispositions, that individuals *have* and that, nevertheless, are episodically experienced only when certain specific circumstances occur – precisely like alarm calls, which are used only when a perceived threat is spotted:

> Although to have an alarm call might be a disposition to emit certain calls when encountering danger, the calls themselves have episodic, not dispositional, natures. […] And likewise, the argument goes, we can agree that what it is to *be* in love with one's partner, or to *be* afraid of heights, or to *be* angry all day for not getting a pay rise, is to be such so as to undergo certain episodic emotions when particular circumstances obtain. So, for instance, if throughout the day Robert doesn't get worked up when reflecting on his failure to get an expected pay rise, then it would seem wrong to describe Robert as being angry that day for failing to get an expected pay rise. However, that is not to say that the emotions themselves have dispositional natures. Indeed, according to the present way of treating these emotional states, the emotions themselves are to be viewed as discrete mental episodes, albeit discrete mental episodes that are undergone when certain circumstances obtain (Whiting 2018: 316 – 317).

For our present purposes, it does not really matter which of the two strategies we embrace. Both views work in avoiding the first issue concerning the existence of unconscious emotions.

But there might be another sense in which an emotion can be said to be unconscious, one that does not refer to enduring dispositions that only episodically show up in the subject's awareness but rather to the actual occurrence of emotional experiences of which the subject is not aware (Hatzimoysis 2007). Some researchers (DeLancey 2002) have linked the occurrence of unconscious emotions in this latter sense to the case of alexithymia. People experiencing this pathological condition are usually incapable of identifying and describing emotions. Moreover, they also seem to be blind to what might have *caused* certain emotional experiences.

Claiming that people with alexithymia undergo unconscious emotions is, at best, a hasty assessment. Indeed, it seems to rely on the highly controversial assumption that in order to be aware of a certain experience (e.g. emotional experience) a subject needs to be able to identify the causes of their emotions or at least capable of conceptually grasping the experience in question by e.g. linguistically labelling the emotion in the right way. We can easily resist both claims. We can imagine that we are experiencing a strong headache that is tormenting us throughout the day. Suppose that it is a Sunday and that we rested all day, we carefully avoided every source of stress and ate healthy meals, we also took our daily portion of physical activity by taking a relaxing walk in the woods. It seems like, in this scenario, all things being equal, we would likely fail to identify any plausible cause for our nightmarish headache. Now we can ask: would this entitle us to conclude that the headache is unconscious? Surely not, the headache is still tormenting us and no matter how hard we try, it remains well above the threshold of consciousness. It is, therefore, utterly unreasonable to describe an experience as unconscious only because we are not capable of accounting for the likely *causes* of that experience.

Similar considerations could be made for the other claim. The fact that we are able to slot an experience into its publicly recognized linguistic category does not allow us to jump to the conclusion that the experience is unconscious. What are we to do with the experiences of infants in their prelinguistic stage of development? Or what about young children who have linguistic skills that are not yet sophisticated enough to properly describe their experiences? Should we conclude that these children are *unconscious* of their experiences altogether?

The case of alexithymia then does not seem to provide a good ground for the claim that emotions can be unconscious. Hatzimoysis brilliantly summarizes the point I have been making so far:

> There are several of our experiences we would find hard to describe, but we do not take this difficulty as a reason for denying their existence. Moreover, we may feel puzzled about what really caused the way we currently feel, but, again, we do not treat this as a ground for denying that we are feeling something. Alexithymia shows how

> sharp these difficulties might become in extreme circumstances – that usually follow upon traumatic experiences that, we may assume the subject would not to want to process conceptually, to describe, and categorize. Conscious experience is not necessarily conceptually structured and linguistically packaged. Therefore, the difficulty of conceptualizing and talking about one's emotional experience provides as such no evidence against the view that one is conscious of the experience. (Hatzimoysis 2007: 395).

We can now move to another case that seems to give more substance to the plausibility of unconscious emotions. It is related to some interesting experiments on the so-called phenomenon of "fear conditioning".[21] In these experiments, subjects, while attending neutral stimuli, were exposed for a very short period of time to a stimulus that was likely to be perceived as fearsome or dangerous. The fear-inducive stimuli could either consist in projected images of things that are likely to be processed as dangerous (e.g. spiders and snakes) or in other generic unpleasant events such as a mild electric shock. The negative stimuli presented were so short that subjects could not be aware of them. Afterwards, when exposed to the neutral stimuli that were previously accompanied by negative stimuli, participants showed some hallmark signs of fear responses such as increased electric skin-conductance or other autonomic changes and bodily alterations. These results have led scientists to conclude that subjects were unconsciously undergoing experiences of fear. Even if the subjects were not *aware* of their emotion of fear, they were nevertheless showing responses that highly correlate with the occurrence of fear.

In response to these very exciting results. I have two remarks to make. The first one is related to the problematic interpretation of these experiments' results as showing the occurrence of unconscious emotions. Indeed, we can interpret these experiments in light of similar considerations to the one I made for the case of alexithymia. We can question again whether not being able to report an experience is enough to conclude that the experience is unconscious.

But there are more substantial reasons to be at least doubtful about the actual occurrence of unconscious fear. As noted by Hatzimoysis (2007) we can claim (i) that the subject is unable "to experience fear in a span of milliseconds for the simple reason that no such experience can occur in a span of milliseconds". But then we would be forced to abandon altogether the claim about the existence of unconscious emotions. Or (ii) we commit ourselves to the controversial view according to which certain autonomic responses and psychological changes (such as increased electric skin conductance) are not only necessary but also *sufficient* conditions for the instantiation of an emotion of fear. I do not have anything against the view that regards certain bodily alterations as a necessary

[21] See (DeLancey 2002.; Ledoux 1984; Zajonc 1984).

component of affective phenomena in general and emotional states in particular, but I think it is untenable to claim that these are *sufficient* for the instantiation of an emotion. Indeed, are we sure it is a sound inference to conclude that subjects are experiencing fear even in the complete absence of other co-occurring distinctive circumstances that centrally characterize the occurrence of fear (i.e. there is no significant behavioral response, the subject is in a safe place, not showing any sign of distress, etc.)? Moreover, the subject might even be well aware of her bodily changes. Still, as already stated, a bodily change *per se* is not sufficient for an emotion to occur: the hair on my skin might rise for a variety of reasons, e.g. someone might have scratched the blackboard.

My final remark brings us back to our talks on reasons and gives us a way to downsize the problem of unconscious emotions for empathy even if we are willing to grant the existence of such satets. Indeed, even if I failed to persuade the reader of the controversy linked to the notion of unconscious emotion, it is still plausible to preserve most of what I said about the emotions as reason-giving states. Actually, my considerations could simply be taken to apply to the case of conscious emotions. Indeed, if empathy is concerned with the imaginative representation of an agent's first-person perspective in a given situation the things of which the subject is not aware, i.e. things that do not show up in her first-person perspective, can be left out of the process. Moreover, since I focused on empathy a tool that aims a providing us with an understanding of the subject's *reason* for acting the way she did, then, again, by definition those elements that are out of the agent's conscious threshold cannot count as her *motivating* reasons for acting, i.e. they cannot count as facts that she could adduce as *speaking for* the way she acted. They would simply be part of that large complex of cognitive processes that, while substantially contributing to the correct functioning of our minds, are nevertheless out of our conscious reach.

6.2. Rationalization as a (useful) fiction?

In a recent paper that appeared in *Behavioral and Brain Sciences*, Fiery Cushman (Cushman 2020) advanced the challenging and fascinating idea that beliefs and desires are often fictional constructs, concocted by agents to produce *post-hoc* rationalizations of their behaviors. The baseline assumption that guides this idea consists in acknowledging the fact that beliefs and desires are not the only drivers of human behavior, not even the main ones. Our behavior is indeed influenced by other processes that are non-rational, unconscious, and yet still biologically adaptive. Cushman groups these other cognitive processes in three distinct categories: instincts, norms and habits. Let us very briefly illustrate each category in order to have a better sense of his claims:

(i) *Instincts*. Instincts are innate dispositions to behave in a certain way when certain environmental circumstances obtain. They are adaptive mechanisms that have been designed through the process of natural selection and that, as such, tend to increase biological fitness. Some of the examples adduced by Cushman are: drinking water when thirsty, fleeing in the presence of a perceived threat, etc.

(ii) *Norms*. They could be seen as the "cultural" equivalent of instincts. Indeed, as natural selection has shaped our built-in instinctual toolkit, cultural selection has made us introject certain norms that have been transmitted through processes of social conformity and social learning. Analogously to instincts, norms selected through cultural evolution tend to be adaptive (Cushman 2020: 6). For instance, Cushman notes how for certain people in the Fiji Islands, it is a taboo to eat certain kinds of fish when pregnant or nursing. Most of these fish are indeed dangerous for the fetus and infants but people do not know it. Their taboo also concerns kinds of fish that are utterly harmless. Therefore, it is safe to conclude that most mothers do not eat those fish just out of a cultural norm. Still, the norm turns out to be adaptive because it actually prevents them from eating harmful food.

(iii) *Habits*. Habits are defined as learned stimulus-behavior mappings. They usually require very little cognitive effort and are quasi-automatic, reflex-like behaviors like turning the light on when entering a dark room or pushing the button of the garage-door opener when approaching our house's garage.

Instincts, norms, and habits pervasively shape our behavior in a tremendous variety of circumstances and, interestingly for our discussion, are *different* from more rational mental attitudes such as beliefs and desires.

It is interesting to notice how Cushman does not explicitly mention emotions in his analysis even if, from what he writes, it is fair to assume that he would slot emotions within the category of instincts. For instance, he refers to cases where we are instinctively disposed to flee when facing a threat, i.e. to a clear manifestation of fear. While I do agree that emotions do certainly have some "instinctive" features – they are definitely, to some extent, the complex result of natural selection – but at the same time it would be a mistake to reduce them to mere instinctive responses, since they are highly sensitive to enculturation processes and do show a great cross-cultural variability (e.g. Lim 2016). Most importantly for our purposes, emotions can provide us with *reasons* for our actions, i.e. they can be legitimately invoked as reason-giving states in their full rights, as I tried to show in the previous sections. Instead, instincts, habits and norms are taken by Cushman as non-rational (even if adaptively

useful) influences on our behavior and, therefore, cannot count, as such, as subjectively held reason-giving states that speak for certain instances of behavior.

But here is the interesting twist in Cushman's paper: even if instincts, norms, and habits are not rational in the sense that they orient behavior independently from any beliefs and desires held by the agent, they can become – and as a matter of fact often *are* – targets of rationalization processes, i.e. they can be re-interpreted through the Humean lens of beliefs and desires during processes of *post-hoc* rationalizations. In a nutshell, behaviors influenced by instincts, norms, and habits can be reinterpreted by the agents performing them – or by other people witnessing certain behaviors brought about by other human agents – by concocting the relevant pair of belief and desire that is (mistakenly) taken to have produced the action.

This is at the basis of what is known in the specialized literature as "representational exchange", i.e. a process through which information stored in one cognitive system is broadcast and made available to other cognitive systems. That is the reason why beliefs and desires are defined by Cushman as useful fictions: on the one hand, they are *fictions* because they are post-hoc constructions that do not mirror the actual psychological processes that contributed to bringing about the behavior in question; on the other hand, they are also *useful* because they broadcast information across different kinds of psychological processes and representations allowing the information-flow to improve the overall adaptiveness of agents in terms of accuracy, efficiency, and flexibility of thinking. For instance, information stored in systems that govern our instinctive behaviors can be transferred to systems that govern reasoning. As a result, valuable information can be treated at different levels of cognition and lead to more adaptive behaviors.

As rightly noted by Cushman, our instincts, norms and habits are storages of "*cached values*", i.e. they store valuable information that is somehow hidden to other cognitive systems. Rationalization routines make those cached values explicitly available to agents and increase their level of fitness by actually allowing them to implement those beliefs and desires when facing analogous circumstances in the future. Information made available to reasoning can be treated at a more abstract and generalizable level with the result of positively impacting e.g. our decision-making and planning abilities:

> A potential function of rationalization, then, is to construct beliefs and desires that are consistent with the adaptive behaviors generated by non-rational processes, and then to adopt them. (Cushman 2019: 5)

A very simple example might clarify this point: we tend to drink water when we are thirsty. This is an instinct and as such the behavior it brings about is not the result of a belief-desire pair held by the agent, but simply by an innate disposition to drink when thirsty. Now, if we translate those instincts into the language of beliefs and desires – i.e. we want water when thirsty and believe that water is worth having around when thirsty – we make this "cached value" (viz. drinking water) available to more rational psychological mechanisms through the process of representational exchange. This can considerably improve our biological fitness which, in this specific case, means increasing our chances to get water when we are thirsty. For instance, we can make plans based on our desire to drink water and belief that water quenches our thirst so that we make sure that we always have access to a sufficient amount of water.

Cushman's brilliant account of rationalization has interesting consequences for what I have been claiming about empathy as a means to understand agents' reasons for actions. Indeed, Cushman's analysis can be extended to the rationalization of both one's own and other people's actions (Cushman 2020: 7). In light of this theory one might reasonably wonder: if beliefs and desires are, at best, a useful fiction, then when empathizing with others in order to understand their *reasons* to act, we always get them wrong. If empathy, as I claimed in the first chapter and pointed out once when discussing Stueber's work, aims at accurately representing the mental process of other people then, when trying to understand agents' reasons for acting in terms of beliefs and desires, empathy fails. As we saw, beliefs and desires are post-hoc constructions that rationalize behavior in ways that are functional to maximize biological fitness. Leaving aside the biological advantages of interpreting other people's behaviors in terms of beliefs and desires, we can turn to the question of accuracy.

First of all, the best way to be charitable to Cushman's account of rationalization is probably to somewhat downscale his claim about beliefs and desires as useful fictions in order to avoid some otherwise manifest inconsistencies within his theory. In fact, if beliefs and desires were *nothing but* the useful constructions that we make up when rationalizing behaviors then it would be impossible to make sense of his subsequent claim about beliefs and desires being adopted by the agent in future instances of analogous behaviors. If there is at least *some* truth in the claim that people hold beliefs and desires after they have been concocted to rationalize an action, then at least in *some* cases people do actually act on the basis of their beliefs and desires. Therefore, empathizing with agents by resorting to beliefs and desires will not only turn out to be useful in Cushman's sense, but also accurate in the sense of actually giving us back a fair representation of the agents' actual process of thought in a given situation.

Throughout the previous sections, I tried to broaden up the domain of reasons by including emotional states as full-fledged reason-giving states, besides beliefs and desires. That move allows me to address some instances of behavior that Cushman would quickly categorize as purely instinctive. As already noticed earlier, fleeing in presence of a threat usually counts as an action grounded in the feeling of fear. In that case, in order to understand the subjects' reason for acting out of fear, the subject would have to factor in the experience of fear in her imaginative representation. Re-enacting fear or, to remain closer to Cushman's terminology, including the state of fear in our post-hoc rationalization routines is what endows subjects with a ready understanding of other agents' behavior.

Extending the domain of reason to encompass emotional states allows us not only to integrate Cushman's analysis of rationalization (as we did with Stueber's account of re-enactive empathy) but also to downscale our concerns about the chances that empathy has of being accurate. For what concerns the first point, the arguments we used to critically assess Stueber's account of re-enactive empathy can be easily extended to Cushman's account of rationalization. I think it is not necessary to go again through the analysis that led us to include emotions as full-fledges reason-giving states. What I think is necessary, though, is to make explicit the implication that such a move would have in the overall economy of Cushman's account: it would require us to categorize emotions as states in their own right, distinguished from both beliefs and desires, on the one hand, and "non-rational" processes such as instincts, on the other. It is certainly the case that emotions are the product of evolution.[22] But rather than being purely instinctual, non-rational processes that shape our behavior they are complex states that compound valuable information in a distinctive affective experience that is likely to exert a readily intelligible influence on our behavior.[23] Most importantly, emotions, by themselves, are sufficient to provide us with the needed information to make readily intelligible certain instances of actions. As insightfully noted by Peter Railton:

> [A]ffect may be a primary source of the information needed to explain oneself and others, or to predict or evaluate what might happen next. (Railton 2020: 31)

In our case it will suffice to stress, once again, the importance of emotions in empathic processes seeking for an understanding of other agents in terms of reasons.

---

[22] Even the leading champion of the so-called theory of constructed emotions, i.e. Lisa Feldman-Barrett (Barrett 2017), who believes that emotions are complex states constructed by the predictive brain on the basis of the subject's previous experiences and broader conceptual competence, would not go as far as to deny that emotions do have an important biological ad evolutionary basis.

[23] I will expand more in detail on what emotions are in the *Appendix* at the end of this chapter.

That is, I believe, both a valuable integration to Cushman's analysis of rationalization and a needed clarification that secures our claims about empathy as understanding of reasons from the threat of being inaccurate by definition.

Onwards

Cushman's paper offers me an opportunity to highlight some important aspects of empathy that will help move the present analysis beyond the topic of reasons. Indeed, if we grant to empathy the possibility of representing through imagination experiential states undergone by other subjects in specific situations, then we can see how the representational power of empathy can be deployed to understand much more than the mere subjectively held reasons for action. By putting ourselves in the other person's shoes, we can come to appreciate her situation from a first-person perspective. This capability to make us face non-actual circumstances grants us the possibility to unlock decision-making mechanisms that would likely take place if we were to actually experience the states of affairs we just imagined.

Indeed, empathy does not always have to explicitly factor-in mental states such as beliefs, desires, and emotions, but can get those states as an output of its imaginative representations. In some sense, by imagining *being* in a certain circumstance, we can get a response that can match the actual response we would have, were we to face that circumstance in real life. As a result of that, empathy, as a form of *simulation* of experiences, can also access what Cushman has called "cached values", i.e. precompiled representations of value generated by previous experiences and that recurrently and irreflexively shape our behavior in our daily life. That is, by simulating experiences, we unlock to some extent some of the (cached) processes that govern our actual behavior in real life, including what Cushman labeled as instincts, norms, and habits.

An example might turn out to be useful. Suppose that instead of knowing already that our friend Jane run away in fear when a huge spider jumped out of an old book she was reading at the library, we are just told that a huge spider jumped out of a book that Jane was consulting at the library and that she ran away from the hall. In the former case, we already know that fear is the main reason for her to act as she did, in the second case we do not know what state made her act the way she did. How do we proceed? Certainly, one possibility consists in trying to imaginatively model the experience of seeing a spider jumping out of a book and work out a likely reaction in response to that. Empathy, as first-person imagining, allows us to do just that. If we are relevantly similar to Jane, our reaction to the imagined representation of a spider is likely to be one of fear. Interestingly enough, though, it is

possible that our fear-response to the imagined representation is unmediated, i.e. we immediately experience a burst of fear (or a fear-like state, for what matters at the moment) in response to what we just imagined. This means that by imagining undergoing experience we can have access to what Cushman called cached values, i.e. we tend to *frame* the imagined situation in a way that is relevantly similar to how we would frame it in real life.

It is easy to see how countless issues immediately raise from the few sentences I just wrote: what are the most remarkable differences between merely imagining an experience and actually undergoing it? Does imagination rely on the same decision-making processes that operate in on-line cognition? What if Jane and I are not relevantly similar? What if our cached value representations are incompatible with hers? What if, due to different enculturation processes or to different psychological dispositions, my norms, instincts and habits are different from Jane's? To what extent can we say that merely imagined and actual experiences are indeed congruent? What are the differences, if any, between imagination and experience that might jeopardize empathy epistemic ambitions?

These are all questions that will be addressed in the next chapters, where I will try to introduce some of the epistemic threats that might muddle empathy's path towards the pursuit of knowledge about other people's minds. Indeed, countless variables do constantly interfere with our attempts to empathize with others. Some of these variables will be the topics of the next chapters in which I will try to delve into the fascinating issue of what empathy cannot do.

Appendix

What is it (like) to imagine an emotion?

Throughout the present chapter, I have been defending the idea of re-enacting emotional states as a way to understand how they can motivate us to act. I focused on the necessity to make room for emotions as full-fledged reason-giving states and on the implications that this move had for empathy as a means towards the understanding of other agents' reasons. That kept me from specifying more in detail how humans are to imagine emotions in the first place. This is no silly question but one that has kept both philosophers and psychologists busy for decades. The present appendix will address this issue head-on in the attempt of making some of the assumptions underlying the arguments presented in the previous sections more solid from both a philosophical and an empirical perspective.

1. Introduction

The occurrence of both, emotion and imagination in tandem, is a landmark of human cognition and yet it is no easy task to give a philosophically and psychologically satisfactory account of their relationship. It is challenging to understand e.g. how it is possible to feel emotions when engaging with merely imaginative representations, react emotionally to the reading of fictions (the so-called "paradox of fiction"), or determine whether imagining an emotion instantiates a fully-fledged emotional occurrence or a mere *fac-simile* emotion.

In the present section, I am going to address some of these concerns from a different angle. Instead of merely asking how is it possible to feel emotions in imagination, or what is the difference between imagined emotions and their real-life counterparts, I am going to take a step back and ask the following, more basic, question: what does it mean to imagine an emotion?

In its simplicity, I believe that this question is well placed to account for several aspects of the link between these two fascinating features of human cognition. It is, in fact, easy to conflate different ways in which it is possible to imagine emotions, and this is likely to generate confusion about what specific issues scholars in the field aim to explain. Different ways of imaginatively engaging with emotions are likely to involve different cognitive systems and have diverging psychological profiles and epistemic outcomes. Being clear about what kind of processes we are interested in is not only a philosophical value in itself but also a necessary ingredient to thoroughly investigate such a richly entangled and hotly debated topic.

Moreover, directing our attention to the various ways in which emotion and imagination can interact will allow me to highlight a surprisingly underexplored form of imagining emotions, which I will call emotion-like imagination (in analogy with belief-like and perception-like imagination). Emotion-like imagination, as we will see, rather than constituting a way of merely imagining *that* a certain emotional experience occurs or a form of *responding emotionally* to imagined representations, requires subjects to actually *re-enact* the emotional state itself so that, as I will put it, they have emotional phenomenal properties as the main content of what is imagined..

After having spelled out in detail what I take this process to be, I am going to offer a plausible empirical ground that I believe could satisfactorily account for the cognitive realizability of emotion-like imagination. As we will see, a promising empirical ground can be found in the emotion regulation literature.

But let us proceed in order.

2. Three ways to conceive of the relationship between imagination and emotion

What does it mean to imagine an emotion? We believe that it means at least three things:

i) Imagining *that* a certain emotion takes place — *Belief-like imagining emotions* (B-Im)
ii) *Responding emotionally* to a certain imagined representation — *Emotional imagination* (E-Im)
iii) Imagining *undergoing* an emotional experience — *Emotion-like imagination* (X-Im)[24]

Let us take a closer look at all of them in succession.

Belief-like imagining emotions.[25] Roughly put, the distinctive feature of this particular kind of imagination consists in imaginatively taking a certain proposition to be true. By means of it, as we already saw in the first chapter, we pretend to believe that the state of affairs described by a proposition actually obtains, i.e. we process it as if it was true.

We can belief-like imagine a remarkably wide variety of propositions such as that Paris is the capital of Spain, that an unknown killer variety of eggplants is spreading throughout Europe, or —to take a

[24] I will use the abbreviation X-Im to avoid confusion with emotional imagination. Moreover, emotion-like imagination, as I will show, centrally focuses on the emotion's *experience* itself, and is therefore a kind of *experiential imagination*, hence *X*-Im.

[25] This is the equivalent of what I called propositional imagination in the previous chapter.

philosophically well-known example— that we are Napoleon. This process endows imaginers with the capacity to explore the vast epistemic landscape of possibility and to draw informed inferences on the basis of merely hypothetical, non-actual, or utterly fictional states of affairs.[26]

There are many issues to be discussed about B-Im.[27] Most of them would lead us too far afield and, therefore, will be left aside. For present purposes, it is crucial to remind us what we anticipated in the first chapter. B-Im can be successfully instantiated without having to represent any of the relevant sensory, affective, and broadly phenomenal components that could be relevantly linked to the propositions imagined. Thus, there is no need to imagine a proposition — as it is usually said in debates about the imagination — *from the inside*, i.e. from the point of view of a subject who first-personally undergoes experiences that might be directly or indirectly related to the proposition imagined. For instance, in order to propositionally imagine that killer eggplants are spreading around Europe, we do not have to model e.g. the outlook of the killer eggplants.[28]

We can easily apply these considerations to the case of imagining emotions. When we are belief-like imagining an emotion, all we have to do is to simply make-believe that a certain emotional occurrence is instantiated. For example, while out on a date that turned out to be disastrous due to our bad mood, we might imagine how it could have been had we happened to feel differently, e.g. happy. On the basis of this counterfactual imaginative assessment, we could conclude that the date would have been much more enjoyable than it actually turned out to be.

Again, even in this case, we can do that without any relevant aspect of happiness being brought about in imagination, we do still feel as miserable as ever. Yet, in virtue of the understanding of the relevant proposition "I am happy" and in virtue of some implicit naïve theory on how people tend to behave

---

[26] Some authors claim that, even though we can imagine a wide variety of propositions to be true, we cannot imagine every proposition we please to be true: there are blocks and limits to what can be imagined. As widely known, debates about such imaginative blocks have tended to gather around the issue of imaginative resistance, i.e. people incapability or unwillingness to imagine certain propositions or states of affairs. See (Gendler and Liao 2016) for an excellent overview. It is far from obvious though that imaginative resistance is something that could arise in cases of mere belief-like imagination. For a sceptical take on this possibility, see (Leeuwen 2016). We will address the issue of imaginative resistance in the next chapter and leave it aside for now.

[27] E.g. how to conceive, in terms of cognitive architecture, of belief-like imagination? How does it relate to other important epistemological and metaphysical issues such as modality and counterfactual thinking? For some excellent overviews on the most stringent philosophical issues linked to imagination see (Amy Kind 2016; Liao and Gendler 2019).

[28] By saying that there is "no need" for concrete experiential features to be instantiated by belief-like imagination, I am not claiming that these features *never* figure in the imaginative process. I am simply stating that they are not essential, even though they do often contribute to the successful instantiation of the imaginative process.

on dates when feeling happy, we can reasonably infer that we would have found ourselves in a much more pleasant situation.

To bring these considerations to their extreme conclusions, we might even hypothesize that someone who has never felt happy in their life could still belief-like imagine to be happy and draw some informed conclusions on the basis of how people tend to behave when happy, even though they do not have the faintest idea of what it feels like to be happy *from the inside*, i.e. even though they are incapable of imaginatively conjuring up any relevant experiential feature of the emotion of happiness.

Emotional Imagination. Another way in which we can conceive of the relationship between imagination and emotion concerns our way of emotionally engaging with certain imaginings. I call this process "emotional imagination" (E-Im).

The principle underlying emotional imagination is fairly simple: we conjure up an imaginative representation of certain objects or states of affairs and, as a result, we undergo some emotional or affective experience in response to what we just imagined. We can, for instance, imagine tasting a dish we always hated and, as a result, coming to feel disgust, or we might imagine bumping into our worst enemy and experiencing a sudden rupture of anger.

One brief clarification is in order here. For E-Im to be instantiated it does not matter whether the emotional responses are caused by belief-like imagination or by other imagery-based or perception-like kinds of imagination. Even if it is more likely that emotional reactions occur in response to an experientially rich representation of objects and states of affairs, I do not have anything to oppose, in principle, to the fact that the mere make-believe propositions can also generate emotional reactions of some sort. What interests me here is that we usually *do* respond emotionally to merely imagined representations, no matter how these imaginings are then represented — as mere make-believe propositions, or as experiences we imagine undergoing from the inside.

It seems from what I said so far that with emotional imagination we never imagine emotional states *per se*, but rather imagine certain objects or events and eventually get an emotional response out of it (Currie and Ravenscroft 2002; Hopkins 2010). As we might put it, emotions seem to be merely incidental to what is imagined, i.e. they tend, at best, to occur as a likely by-product of what we imagined.

Even if we might know with a high degree of confidence that certain imaginings will bring about certain emotional responses, we get the relevant emotion only in virtue of statistical regularities in our affective responses. In a nutshell, with emotional imagination we can never have the emotion as

the main *content* of what it is imagined, but always have to work it out indirectly by imagining something else.

These considerations help me introduce the third way in which it is possible to conceive of the relationship between imagination and emotion.

Emotion-like Imagination. We can now ask: Is it possible to imagine *from the inside* undergoing an emotional experience without having to work it out indirectly by imagining something else that is likely to trigger certain emotional reactions?

To answer those questions, we have to wrap our heads around what it would mean to experientially imagining an emotion or to emotion-like imagine as opposed to the other kinds of imagining emotions that we reviewed so far.

Emotion-like imagining anger consists in neither belief-like imagining *that* we are angry nor in imagining an anger-inducing object or state of affairs and, then, get anger as a response to what we imagined. It rather consists in imagining *feeling angry*, so that we can appraise some objects or facts *through that specific emotional lens*. In this sense, the subject would have to imagine being in a certain state that feels relevantly similar to the emotion they are trying to recreate and that will eventually constitute the basis for an emotional appraisal of the specific intentional object toward which it is directed.

I believe that the phenomenon of experientially imagining emotions or emotion-like imagination (henceforth, X-Im) has been somewhat neglected in the specialized literature, which has tended to privilege the other two kinds of imagining emotions I have mentioned just above. Indeed, scholars (especially in the field of philosophy) have often circumscribed their concerns for imagination and emotion mainly within the perimeter of the paradox of fiction and the tightly related issue of the ontology of emotions (e.g. Currie 2020; Currie and Ravenscroft 2002; Gendler and Kovakovich 2006; Walton 1990). This has implicitly kept the hostage of E-Im, since the kinds of example they mostly engage in involve cases where a subject responds emotionally to an imagined situation, as it paradigmatically happens when reading works of fiction.

A thorough account of X-Im is, therefore, not easily found in the literature. An exception to this widespread tendency is offered by Goldman (2006). He indeed posed the question concerning X-Im in a way that is very similar to mine:

> I can imagine seeing a yellow parrot, feeling sad, feeling outraged, or feeling elated. It is also possible, no doubt, to imagine that one feels elated, which is equivalent to assuming the truth of the proposition ''I am elated.'' But there is another way to imagine feeling elated, namely, to conjure up a state that feels, phenomenologically, rather like a trace or tincture of elation. […] When I imagine feeling elated, I do not merely suppose that I am elated; rather, I enact, or try to enact, elation itself (Goldman 2006: 47)

However, his remarks are rather diluted in the much broader context of making his case for simulation theory, which he is best-known (and criticized) for. Moreover, a large number of his further remarks about emotions and imagination mainly fall within the category of what I called E-Im (e.g. Goldman 2006a), [29] so that we are still left without a detailed account of how X-Im of emotions is supposed to work.

Throughout the rest of this appendix, I am going to explore more in detail the phenomenon of X-Im and offer a plausible story regarding its eventual cognitive realizability.

3. Some preliminary concerns

At this point, I believe that it might be useful to take care of four distinct worries one might have at this stage of argumentation. The first (i) regards the distinctions I drew between various ways of interaction between imagination and emotion, the second (ii) X-Im's conceivability, the third (iii) its actual existence, the fourth (iv) its usefulness. Let me briefly address all of them in turn.

(i) For what concerns the first worry, the distinctions I drew might look too clear-cut, i.e. they seem to tell apart processes that, *de facto*, are strictly intertwined and rarely, if ever, occur in isolation from one another. This is an ancient problem in the mind sciences, already noted by William James when he wrote that: "Consciousness [or the mind as we might put it] does not appear to itself chopped up in bits" (James 1890: 239).

    In describing the mind, we often fall prey to the mistake of drawing distinctions where, in fact, entrenchment is the rule. This observation rings true also in the case of imagination and emotion, where the differences in place might look too subtle to matter. However, I do believe that, even if in the context of everyday cognition the processes enlisted above rarely occur in isolation, they present differences that, however subtle, are nevertheless remarkable and worth being pinned down. Being aware of them is not only a value in itself in the pursuit of a worthwhile description of the processes involved, but also provides us with powerful

[29] I.e. emotions as outputs of simulated states of affairs we imagine to experience.

conceptual tools that allow us to highlight with greater clarity issues of philosophical and psychological importance.

(ii) For what concerns the second worry, one might claim that emotions *always* have an intentional object. In this sense, it is impossible to merely imagine an instance of anger, without also imagining some kind of intentional object, no matter how vaguely defined it is. In this sense, one might not see how X-Im might even be conceivable. Since imagining emotions will always be imagining feeling emotional *about* something, X-Im could boil down to mere instances of emotional imagination.

I believe this worry is misplaced. Of course, our imaginings would have to minimally contextualize emotions so that they be *about* something. Still, this does not entail that X-Im boils down to E-Im. There is nothing that forces us to think that imagining the object necessarily predates and generates the emotion as its side-effect. Indeed, it looks perfectly conceivable to me a process in which a certain emotional state constitutes the main content of what is imagined, even if that would require us to preserve its necessary intentional structure, i.e. even if we will always have to imagine it as referring to *something*. I will come back to this issue more extensively in the following sections.

(iii) We can now turn to our third worry. Indeed, even if we grant that something like X-Im is conceivable, we can still be concerned by the issue about its actual existence, and ask the following more stringent questions: can X-Im concretely find its place among the other cognitive processes in our minds? That is, can we actually imagine emotional states in a way that is relevantly similar to other states we can imaginatively recreate such as perceptions? Currie and Ravenscroft (2002) count among the most well-known theorists in the field who have given a negative answer to these questions. Even if they are willing to grant to imagination a fair amount of recreative power, i.e. they believe that our mind is populated with imaginative counterparts of a wide variety of non-imaginative mental states (e.g. belief/belief-like imagining, desire/desire-like imagining, perception/perception-like imagining), they deny that emotions can be reproduced within the space of imagination. They argue that emotions are indeed "transparent to imagination" (Currie & Ravenscroft 2002: 190), i.e. that imaginative representations bring about full-fledged emotional states, rather than mere imaginative counterparts of them. In this sense, it is impossible to locate emotional states within the perimeter of what can be imagined: emotions can only be actually undergone and never merely imagined. They offer two main reason to embrace their view: one

phenomenological and one evolutionary. The phenomenological reason relies on the introspective evidence that we actually undergo emotional experiences when we entertain imaginative representations. For instance, if we imagine something amusing, we can respond with *actual* amusement to it and not with a mere pretend version of it:

> There is no imagining that has an amusement-like character; there is only being really amused. […] As part of an imaginative project, one is often actually amused by some aspect of what is imagined, and here one's amusement is caused by what one imagines, as well as having the imagined event as its intentional object. In that case we can say that imagination is transparent amusement. (Currie & Ravenscroft 2002: 190)

The evolutionary explanation relies on the considerable advantages deriving from a cognitive system that is able to respond emotionally to what is imagined. Indeed, being able to emotionally assess the imaginative representations we happen to entertain offers huge benefits to our capacity for decision-making and planning. When wondering whether we should take the dark, creepy road back home or the well-lit, comforting one, a fear response towards the first imagining is what might providentially prevent us to run into unnecessary troubles on our way back home. As they put it:

> [A]n emotional sensitivity to merely imagined circumstances can help me manage my affairs: in order to affect my predicament I must act; to act effectively, I must plan; to plan, I must imagine alternative scenarios and choose between them. Having a system of emotional responses poised to respond to what I imagine is a capacity we would expect to find in creatures able to choose between alternatives. (Currie & Ravenscroft 2002: 197)

To these lines of argumentation, I do not have anything substantial to oppose. I indeed happen to agree with most of what Currie & Ravenscroft have to say about the interaction between emotion and imagination. It is what they *do not* say about these issues that I want to address. They claim that there cannot be emotion-like imaginings but, in fact, they only give arguments in favor of the actual occurrence of emotional states in reaction to imagined representations: the cases discussed by the two authors fall within the category of E-Im, i.e. they all concern cases in which we *respond emotionally* to imaginings. From this angle, they conclude that imaginings can generate full-fledged emotional responses on the basis of their persuasive phenomenological and evolutionary considerations. But, then, there is still a story to be told about imagining emotions in the relevant sense I tried to capture through the notion of X-Im: what about targeting emotional phenomenal properties as the main *content* of what is imagined, rather than having them as a mere response to what we imagined? What about imagining what it feels like to undergo a certain emotional state? The authors remain largely

silent on these questions. Indeed, even if it is true – as I also happen to believe – that we can have genuine emotional responses to what we imagined, this by itself does not rule out the possibility that we can target emotional phenomenal properties (i.e. how emotions *feel* like) as the main content of what is imagined, i.e. it does not preclude the possibility of something like X-Im.

As it will become evident throughout the rest of this appendix, the present account of X-Im will reveal some affinities with Currie & Ravenscroft's considerations about the transparency of emotions to imagination. As we will see, X-Im of emotions will actually re-instantiate some of the relevant phenomenological properties of full-fledged emotional states. At the same time, they could still be counted as imaginative states since X-Im voluntarily generates and instrumentally co-opts emotional phenomenal properties to fill and sustain its imaginative enterprise. I will argue that this goal can be achieved thanks to emotion regulation strategies. More on that in the next sections.

(iv) For what concerns our fourth and final worry, I believe that one of the most valuable and interesting uses of X-Im emotions consists in its capability of enabling subjects to imagine emotional perspectives that considerably differ from their own. Suppose I happen to love dogs: whenever I see a dog, I am overwhelmed by feelings of joy and amusement. As an inveterate dog-lover, I might find it hard to imagine the perspective of someone who finds them disgusting. How could I ever understand the perspective of a dog-hater *from the inside*? I believe that X-Im could come to rescue. Being able to imaginatively re-enact some relevant aspects of the emotion of disgust could help us restructure our cognitive and attentional resources in a way that could make salient and intelligible to us some features of dogs that make them suitable for an instance of disgust. X-Im would help us reconfigure our perspective so that we can see the world through the eyes of somebody who relevantly differs from us.

In light of what said above, it comes naturally to realize the crucial role that X-Im might play in the empathic re-enactment of emotional states.

Having taken care of those preliminary concerns that could arise at this stage, I shall now build my case for X-Im of emotions.

4. Substantiating emotion-like imagination

So far, I have been distinguishing different ways in which imagination and emotion can interact: B-Im, E-Im, and X-Im. I have also argued that the latter, X-Im, has not received the deserved amount of attention in the specialized literature. In the present section I start to partially fill this gap, I will

try to illustrate more in detail what X-Im is and how it can be more in detail distinguished by other ways of imagining emotions. This has to be taken as a first step in an area that is yet to be adequately explored.

As it should be clear by now, X-Im is the process through which we imagine from the inside to undergo a certain emotional state: far from constituting a mere detached intellectual exercise (like e.g. B-Im) X-Im requires us to *re-enact* how the emotion feels from a first-person perspective. As already anticipated, by re-enacting an emotion we come, eventually, to undergo an actual emotional experience, i.e. through X-Im we enter in a state that relevantly resembles what it feels like to undergo the emotion in question.

Getting clearer on some of the essential features of emotional experiences will help me highlight more in detail the concrete workings of X-Im. In what follows I will enlist some of the central features that are usually found in the specialized literature on emotions and that, therefore, can be counted with a sufficiently high degree of confidence among the distinctive features of affective experiences in general and emotions in particular.

- *Phenomenal valence*. One distinctive feature of emotional experience is the so-called phenomenal valence, i.e. the felt positivity/negativity of certain experiences (Charland 2005; Teroni 2018). In the literature the positivity/negativity feature can be further elucidated in hedonic terms as pleasantness unpleasantness or in value terms as seeming value or disvalue (Carruthers 2018). It should be emphasized that valance is a *phenomenal* property of emotional experiences, i.e. something that is consciously experienced by the subject in an emotional state.

- *Bodily sensations*. Another prominent feature of emotional experiences are bodily sensations. Emotional experiences are paradigmatically accompanied by changes in muscle activity, breath, or heartbeat. A particularly noteworthy feature of this somatic phenomenology is *felt arousal*: during an emotional experience the subject feels a more or less localized increase/decrease in the level of energy or excitement (Colombetti and Harrison 2018). Again, with bodily sensations and felt arousal, it should be stressed that, in this context, we are interested in their phenomenal dimensions.

- *Motivation*. The aforementioned features of emotional experiences are usually associated with the ability of emotional experiences to motivate us to act or behave in a certain way (Corns 2014;

Scarantino 2014). Thus, emotional experiences can be said to be *motivational*, i.e. they can exert a motivational push that can be directly translated into certain behaviors or behavioral tendencies.

The idea behind X-Im is that, in order to imagining undergoing an emotional experience from the inside, we would need to be able to conjure up the emotional phenomenal properties I just mentioned. In other words, the imagined content distinctive of X-Im are said phenomenal emotional properties. This is what ultimately allows us to significantly distinguish E-Im from X-Im: the former kind of imagination, indeed, consists of imaginings whose content are objects or states of affairs that are likely to evoke certain emotional occurrences but are not intrinsically emotional themselves. Interestingly, E-Im is qualified only in virtue of *what* it is imagined by rather in virtue of *the consequences* brought about by certain imaginings. In other words, E-Im could be said to be the cognitive-basis of the actual emotions. That implies that one could, in principle, imagine *anything* and that could still count as an instance of E-Im, as far as a relevant emotional experience is brought about.

This contrasts with X-Im, whose imaginative content cannot simply be anything, but mainly emotional phenomenal properties. In a nutshell, differently from E-Im, X-Im is constitutively defined by the content of what is imagined.

Now, it is rather implausible that X-Im could really get off ground, were it to consist in *merely* imagining the aforementioned emotional phenomenal properties (valence, arousal, bodily sensations, motivation). Indeed, X-Im can be significantly facilitated by another central aspect of emotional experience that has not been mentioned so far: intentionality.

As already anticipated in the previous section, paradigmatic emotional experiences are intentional states (e.g. Goldie 2002; Kriegel 2014), i.e. they are about something, they represent states of affairs as being a certain way. My experience of fear represents the spider as being dangerous and my experience of disgust represents the smell of a skunk as repellent. These properties are called the *formal objects* of emotions (Kenny 1963; Teroni 2007): they are basically the emotion-specific properties that are ascribed to the emotion's specific intentional objects. So, to sum up, we can say that emotions have two intentional objects: the particular object or state of affairs towards which it is directed and the formal object which is represented as borne by the former.

In light of what has just been said, the intentionality of emotional experiences relies on other mental states such as perceptions, beliefs, memories, etc. These mental states provide the emotions with the particular representational content or *base* for the emotion (Deonna and Teroni 2012). Similarly, as

for actual emotional states, outputs of X-Im need also to rely on a specific *base* in order to achieve their representational goal: they are in need of an intentional anchor for the experiential imagination of an emotion. Paradigmatically, these intentional anchors are going to be other kinds of imaginings, but it does not have to be necessarily that way. I can indeed imagine being scared of squirrels by simply having a look at the one I am seeing right now on the tree in front of my window: in this case, my actual visual experience of a squirrel is what can constitute the intentional base for my X-Im of fear (of squirrels). Usually though, it seems like the intentional base of X-Im will be worked out by other kinds of imaginings that will be conjured up in order to bring about the desired imaginative representational output.

These considerations do not entail that the intentional elements necessarily have to predate the imagined emotional phenomenal qualities. There is room for the latter to co-exist with or even exist without the former. However, it is still perfectly plausible that, in bringing about a certain X-Im output, we *first* evoke the intentional base and then try to work out the desired emotional experience. Namely, if I am attempting to imagine being disgusted by dogs, I can *first* conjure up the relevant image of a dog. Note that it might well be – as in fact is – that my spontaneous reaction to an image of a dog is not one of disgust but, rather, of joy and amusement. In a second moment, I can try to detach the image of a dog from the spontaneous emotional reaction I tend to have and try to couple it with one of disgust by evoking the relevant emotional phenomenal properties of that emotion.

It remains an open question now whether X-Im could have an extended lifespan without being directed towards an intentional object. Emotional phenomenal properties might be brought into existence by X-Im but not really sustained for long without something they are about. Thus, even though emotional phenomenal properties can be conjured up by X-Im, it is questionable whether they are able to significantly evolve. Indeed, if we try to imagine a phenomenally valanced state, then we have to imagine entering a state that feels positive or negative: but emotional experiences do not usually feel positive or negative *simpliciter*. Emotions feel positive or negative in relation to their formal objects, e.g. amusement feels positive in relation to funniness and fear feels negative in relation to danger, and these formal objects are represented as borne by the specific intentional objects emotional states are about. Therefore, it surely appears hard to evoke the required phenomenology of emotional states without also anchoring it to an intentional base.

With these resources in place, we can, once more and more efficiently, highlight the specific features that tell X-Im apart from other ways of imagining emotions. Let us imagine Napoleon after his defeat at Waterloo. We can B-Im *that* Napoleon was furious while contemplating the grim spectacle of his

military debacle. We can even enrich our belief-like imaginings with images of the devastated battlefield and of the restless and upset expression on Napoleon's face. Now, we can imagine that, due to our specific personality and psychological dispositions, we might *react* in a certain way to these imaginings, say with an emotion of *Schadenfreude* for Napoleon's anger. This case would count as E-Im. But, in fact, we could react to these imaginings with an open-ended variety of emotions, they would all still count as cases of E-Im. Said it differently, with E-Im there is no *mapping relation* between what is imagined and the emotions evoked: what is needed is simply an emotional reaction of some sort, whatever it may be. With X-Im, instead, there is a mapping relationship between the emotions evoked and the imaging that is brought about. If I try to imagine *being* the angry Napoleon, I have to bring about the relevant emotional properties that might help me to shape the rich phenomenology of his imagined emotional state. X-Im can be said to be successfully instantiated only if the emotion evoked is the one the subject is trying to experientially imagine.

But then, what does it mean to conjure up the required emotional phenomenal properties? As already anticipated in the previous section when discussing Ravenscroft and Currie's account of imagination of emotions, there is a relevant sense in which we can say that emotions are largely transparent to imagination. This implies that, in order to evoke the phenomenal qualities that are constitutive of emotional experience, the subject needs to enter a state that phenomenologically resembles the *bona fide* emotional states. Said otherwise, imagining emotional states from the inside, brings about an emotional state itself. In this respect, X-Im is still very similar to E-Im because it too results in us undergoing an emotional state. And yet, there is one, crucial, difference: E-Im is *defined* by us entering an emotional state, whereas in X-Im we enter an emotional state only if it is successful. Moreover, as already stated previously, X-Im has specific emotional *targets*, whereas E-Im has just emotional *consequences* that are not constitutive of what is imagined and only incidentally related to what is imagined.

Now, one might wonder, how can X-Im be told apart by usual garden-variety emotional states? Moreover, does the fact that X-Im's outputs are constituted by actual emotional phenomenal properties, i.e. by states that are phenomenally largely equivalent to the actual experience of emotions, *per se* rule out the possibility for X-Im's to count as an imaginative process?

First of all, in contrast to voluntary actions, emotional experiences are not something that we usually decide to bring about: they seem to simply happen to us. In a nutshell, we are not emotional *agents*, but rather emotional *patients*. In the case of X-Im, the emotional states that are brought about are

outputs of some kind of voluntary mental action.[30] We decide to imagine from the inside the phenomenology of being in a certain emotional state and we get, as a result, an experience that feels relevantly similar to the full-fledged emotional state we are trying to represent.

From this, one should not conclude that every instance of X-Im counts as voluntary in the sense that the imaginer should always be fully aware of its epistemic goals and interests. It might well be the case that sometimes imaginers simply *find* themselves imagining certain emotions from the inside, rather than consciously deliberate about imagining them. Still, this does not preclude the fact that X-Im is something that someone *does*, i.e. some kind of (mental) action we bring into existence and that is still subject to the will "in the sense that it makes sense to direct the will at them and that the will can influence their course and character" (McGinn 2005: 14).

Now, one might still legitimately claim that *bona fide* emotional states as well are, at least to some extent, subjected to the will, in the sense that we can sometimes voluntarily shape and influence their course: we can intend to calm down, we can take action to "defeat our fears", etc. These are all instances where our will does play a role in shaping and regulating the course of our emotional experiences. Yet, it is still intuitively evident that there is a difference between bringing into existence an emotional state with a representational intent (implicit or explicit for what matters), as it happens with X-Im, and merely influencing or regulating a pre-existing emotional state by means of the will, as it happens with more central instances of emotional experiences. In the former case the will seems to play not only a *regulative* role but also a *generative* role in the sense that there it is the subject's mental action that is responsible for the emergence of the emotional state in question. Said otherwise: the emotional states we find ourselves in when experientially imagining emotions do not depend on external stimulus but are rather *endogenously* produced by the imaginer.

McGinn made some insightful observations when contrasting images (i.e. the outputs of imagination) and percepts:

> Subjection to the will implies a difference in the *causation* of images and percepts. Percepts typically have their causal origin in external stimuli; and even in the case of total hallucination, the causation does not involve the subject's decision-making mechanisms […]. But in the case of images, the causation comes from "inside"; it is endogenous not exogenous. It is very tempting to resort to the idiom of agent causation: it is *I* who cause the mental images that occupy my consciousness. (McGinn 2005: 15)

[30] More on this aspect in the next section.

These considerations can be easily extended to the case of emotional experiences and X-Im. It is indeed the imaginer who is causally efficient in evoking the emotional experience in question.

But then, a sceptic of the present proposal might wonder, similar considerations could be made for E-Im. After all, the imaginative representations that are at the base of E-Im are also undeniably produced by the subject and to some extent subject to their will. To weight the actual impact of such a critique one should be careful about what the imaginer is actually doing in X-Im and E-Im outputs respectively. In fact, while X-Im is an intentional imaginative act in itself, E-Im as such, is not an intentional act at all. We have seen that E-Im can be the output of an intentional act, however. You might e.g. have the intention to cheer yourself up by envisioning that a friend will visit you soon or enter the door and smile at you. Now, you might respond with joy to this intentional imagining. If so, then this would qualify as an E-Im that is brought about by an intentional act. However, imagining your friend might just as well fail to evoke an emotion in you and so there would be no E-Im. Note that the same intentional act might actually lead you to feel sadness (instead of the intended joy) because the latent fact that you miss your friend becomes salient upon imagining his entrance. In this case we would still speak of E-Im despite the fact that sadness was not the intended emotion.

In a nutshell: with E-Im the imaginer is actually exerting control on the imaginative representations but not (directly) on the kind of emotional reactions they experience in response to them. Indeed, the imaginings themselves are responsible for what the imaginer is feeling, i.e. the imaginer is an agent with respect to what is imagined but a patient with respect to what is felt. In the case of (successful) X-Im, instead, the imaginer is exerting control on the emotional state itself, i.e. they are an agent with respect to what is felt.

Finally, X-Im's outputs, can still count as imaginative even if they rely on the actual re-experiencing of some the relevant emotional phenomenal properties we aim to evoke. Indeed, this should be simply taken as the way in which the imaginative representation is concretely brought about in our cognitive architecture. The fact that, in the attempt of imaginatively representing the experiential dimension of emotional states, we undergo (if we are successful) an emotional experience that feels relevantly similar to the one we are trying to imagine, does not imply that X-Im is not an imaginative act, but simply that the process of conjuring up emotional phenomenal properties is partly realized with the contribution of the same properties we are trying to imaginatively evoke.

5. From emotion regulation to experiential imagination

Having clarified some of the essential features that I take to be instantiated by X-Im, in the present section, I am going to provide some empirical ground for the claims that I have been putting forward so far. Namely, I am going to offer some elucidations about how a process with the features I ascribed to X-Im can be concretely realized by our minds.

Luckily for us, X-Im is not the only case in which individuals exert an indirect influence on their emotions. The common phenomena of influencing one's emotion has been widely researched under the header of Emotion Regulation (ER). Although no ER studies to date have directly addressed the process of X-Im, it is possible to build on existing research to postulate empirically plausible X-Im processes.

Emotion Regulation research emerged as a distinct domain in the 1990's (Campos et al. 1989; "Emotion Regulation and Mental Health - Gross - 1995 - Clinical Psychology: Science and Practice - Wiley Online Library" n.d.) and has witnessed a vast surge in output since then (for a recent overview, see McRae and Gross 2020). As a result of this vastness, the ER research literature is fairly heterogeneous. There is a large variety in both the strategies that people use to influence their own emotions (for a taxonomy, see Koole 2009) and in affective styles (Hofmann and Kashdan 2010), as well as a large difference in ER ability from one individual to another (Gratz and Roemer, 2004), which results in different ER models. Nevertheless, since the relationship between ER ability and ER strategy is bidirectional, different ER models should be regarded as complementary rather than exclusive (Tull and Aldao 2015).

The most prominent model of strategies of ER is Gross's process model (Gross 1998, 2015). The process model will provide us with a useful organisational schema to understand ER with the subsequent aim of characterising X-Im[31]. The process model considers emotion generation to be a process in which the individual encounters a situation, attends to the relevant aspects of the situation, appraises it in relation to active goals and has a series of responses (experiential, physiological and

[31] Here, I do not commit to the process model wholesale. In particular, there are some worries that the process model might give us an overtly compartmentalised view of emotion generation and regulation. On this issue, Barrett and colleagues argue that "emotions are not unique mental states that are caused by dedicated mechanisms, to be modified by another set of dedicated regulatory mechanisms. Instead, emotions emerge, and regulation occurs, as the consequence of an ongoing, continually modified constructive process that makes sensory inputs meaningful." (Barrett et al. 2014: 448). I agree with this critique, and I simply use the process model as an organisational schema that is useful to identify relevant families of ER strategies, and to highlight the importance of target states in some instances of ER. X-Im, as I will go on to argue, is one of those instances.

behavioural). Five families of ER strategies are then identified in relation to the emotion generation stage at which they first intervene:

- situation selection (e.g. declining to watch dramas in order not to get sad),
- situation modification (e.g. putting on calming music to avoid getting anxious),
- attentional deployment (e.g. distracting oneself away from emotionally charged stimuli),
- cognitive change (e.g. trying to take a more equanimous attitude towards an anger-inducing event)
- and response modulation (e.g. slowing one's breathing to calm down).

According to this conception of ER, "the defining feature of emotion regulation is the activation of a goal to influence the emotion trajectory" (Gross 2015: 3).

There is one main obstacle that runs against the project of applying ER to the case of X-Im: we might call it the *generation issue*. This issue stems from the fact that there seems to be a substantial difference between standard instances of ER and X-Im because, as already anticipated in the previous section, when assessing the issue of how we can influence and shape the course and profile of our emotional experiences, ER research has tended to focus on the reduction of intensity of pre-existing negative emotions, whereas X-Im requires the *generation* of specific instances of emotions to fulfil the imaginative enterprise.

Indeed, the large majority of experiments have studied ER based on hedonic considerations. The general assumption in many experimental settings is that subjects regulate emotion in order to increase pleasure and decrease displeasure. In other words, the general assumption is that the target of ER is *modulating valence* rather than changing or generating emotional states.

However, recent research has shown that subjects can also engage in ER out of utilitarian considerations (M. Tamir et al. 2008). This research shows that individuals seek to experience emotions that are considered useful even if they are unpleasant (e.g. they try to become worried before facing a potential threat). It is no surprise that studies researching utilitarian considerations in ER have found that subjects aim to elicit but also target emotional states, such as happiness when expecting collaboration and anger when expecting confrontation (M. Tamir et al. 2008; M. Tamir and Ford 2012).

The utilitarian-considerations experimental work is in line with the process model of ER introduced earlier according to which ER involves crafting an emotion trajectory with the aim of reaching a goal.

In some cases, said goal is itself a target emotion state[32], and the trajectory is the process that leads the individual to it. X-Im is one such case. X-Im turns the standard process of emotion generation (as characterised by Gross's process model) around. In normal circumstances, a situation causes an emotion through appraisal, attention patterns and responses. In light of existing ER research, it seems fair to postulate that during X-Im, a reverse engineering process takes place, in which all of the other elements of emotion generation are transformed and modulated through a combination of ER strategies until the target emotion emerges.

Two good examples of the modulation of physiological responses to elicit emotions that can be extrapolated to X-Im are breathing patterns and strategic muscle tension and relaxation. Specific breathing patters are associated with different moods and emotions (Boiten et al. 1994), so they provide a clear way in which to elicit those emotions (if a subject wants to imagine anger, heavy breathing is a good way to go).

Most research on muscle activation and ER is naturally about muscle relaxation, rather than about muscle tension (Esch et al. 2003). Much of this work builds on Jacobson's technique in which individuals tense and relax different muscles of their body (Jacobson 1928), which can be used to alleviate anxiety (Rausch et al. 2006). Nevertheless, it stands to reason that the same principles could be used to elicit a highly arousing emotion by tensing certain muscles during X-Im. What is interesting is that muscle relaxation techniques are most effective when combined with attention deployment strategies, such as paying attention to muscle sensations (Borkovec and Hennings 1978). This, again, supports the idea that X-Im is most likely to combine several ER strategies holistically.

Another physiological response strategy that has not been studied extensively in the ER literature could be the modulation of facial muscles (Davies et al. 2016) A wrinkled nose leads to experiences of disgust, a wrinkled forehead to negatively valenced experiences and raised eyebrows to an experience of surprise (M. B. Lewis 2012). Such facial responses could be modulated in order to elicit target emotion states in X-Im.

The existing empirical evidence points to the modulation of physiological responses as the bedrock of X-Im. Bodily movements alone might be able to activate emotional experiences (Niedenthal et al. 2005). This means that physiological modulation on its own might be enough for some instances of

[32] Of course, in the case of X-Im it is likely that there is a larger overarching goal, such as taking someone else's perspective.

X-Im (e.g. heavy breathing, frowned eyebrows and tense shoulders might suffice for the subject to experientially imagine being angry).

At this stage, a central question still awaits to be answered in light of the considerations that have been put forward in the present section: could there be then instances of solely affective X-Im that does not involve an intentional object? This links with the question raised in the previous section of whether or not X-Im could be successful without invoking intentional objects. A stint of introspection indicates that eliciting an experience of valence in such a way seems to be possible, at least for some individuals (e.g. try feeling negative valence without eliciting visual imagery or directing it to surrounding objects — it is difficult and the resulting valence is weak, but it seems to be possible nonetheless). However, as also stated in the previous section, it is doubtful that a high intensity, clearly defined instance of X-Im could arise without other ER strategies involving imagery, cognitive reappraisal, etc. The case is even harder to hold for more complex emotions, such as mistrust or nostalgia. Hence, the modulation of physiological responses can serve as a form of emotional *generation*, but a full-blown emotional state requires the scaffolding of other types of ER strategies that serve as a form of what we might call emotional *inflation*.

Going back to a previous example, one might be able to evoke a tinge of anger with physiological modulation alone, but to experience the feeling of being the angry Napoleon upon his defeat at Waterloo, a further scaffolding of imagination will be needed. Furthermore, the iterations of different strategies will loop, feeding back on each other to elicit, intensify and mould the target emotion. To imagine such a precise emotion as the anger of a defeated Napoleon, we might change our posture, hunch, pay attention to and enhance bodily sensations of exhaustion and frustration, imagine the rain and the mud of Waterloo, imagine the words of French officials bringing news of the defeat, reappraise the imagined situation to become angry at Wellington and even at the rain, react then to this vivid scene with new physiological responses, making our breathing heavier and furrowing our brow, and so on and so forth in a rich, iterative poly-regulated process aimed at imaginatively enacting a very particular emotion.

Onwards: Back to empathy

With the present appendix I hope to have accomplished at least three things: i) offering a useful typology of the different ways in which imagination and emotion can interact, ii) unveiling a rather neglected way of imagining emotions, namely X-Im, iii) providing a sound conceptual scaffold and a plausible empirical basis for the process of X-Im.

I am confident that a clear typology of emotion and imagination is a useful conceptual tool that could allow scholars in the field to successfully keep track of what people are talking about when addressing the issue of imagination and emotion.

Moreover, this typology made us mindful of a particular way of imagining emotions, i.e. X-Im, which has an intriguing psychological profile and exciting potential epistemic outcomes. I hope at least to have paved the way for further exploration of a rather underexplored topic.

But most importantly, for the sake of the present work, I believe X-Im clarifies some important aspects of empathy and the way in which it operates in providing insight into the experiential dimension of other people's minds. More precisely, X-Im allows me to clarify a way in which it is possible to achieve phenomenal insight into emotional states, which is what we needed in the first place to account for the possibility of re-enactive empathy for emotional states. Furthermore, as already announced at the beginning of the present appendix, when deployed in tandem with other imaginative strategies, X-Im can turn out to be an invaluably powerful tool to grasp emotional perspectives that significantly differ from ours and overcome some structural shortcomings of E-Im. Indeed, when trying to imagine a situation as experienced by another person who has an "emotional take" on it that is incompatible with that *we* would have when facing the same circumstances, then X-Im can come to rescue. It could help us enter an emotional perspective that we would not spontaneously occupy and, by allowing this, it would provide us with increased chances of understanding different others.

Of course, X-Im cannot always succeed and, most importantly, cannot always be enough. It is common, indeed, that we fail to enter perspectives that would bring us too far away from the individuals we are, i.e. too far away from the set of preferences, beliefs, values we hold, so that it becomes impossible for us to empathize with people (or perspectives) that are drastically different from us (or ours).

These issues will constitute the main concerns of the following section, which will explore the territory of empathic failures and limits. After having explored the epistemic possibilities of empathy, i.e. what empathy *can* do, it is now time to turn to the issue of what empathy *cannot* do.

Chapter III

What Empathy Cannot Do (I): The Bounded Self

In this chapter, I am going to tackle a bundle of issues that are likely to play a major role in what we may call "empathic failures", i.e. those instances of empathy that simply miss their goal and fail to deliver insight into other people's perspectives. This family of issues can be gathered under what I will call the problem of the *bounded self*. By this expression I mean that the empathic imagination is ultimately bound to the kind of individuals we are. Roughly put: even if we are capable of imagining a remarkably rich array of non-actual things and states of affairs, what we can imagine from a first-person perspective is ultimately *bound* to the persons we are, i.e. to our preferences, values, commitments, etc. This is what anchors and ultimately grounds our imaginative self in our actual self, therefore preventing the former to "migrate" too far away from the latter. Of course, the kind of subjects we are is also what allows us to model other non-actual subjective perspectives that we do not currently occupy: it allows us to imagine experientially vivid perspectives. That is, who we are plays a pivotal role in both allowing us to empathically understand others and setting the limits of our empathetic capabilities.

1. Introduction: What Pessoa got wrong about the imagination

In a beautiful passage of his *Book of Disquiet*, the Portuguese writer Fernando Pessoa explored the nature of imagination, specifically of what it takes to imagine being someone else. He wrote:

> "I can imagine that I'm everything because I'm nothing. If I were something, I wouldn't be able to imagine. An assistant bookkeeper can dream he is the Roman emperor, but the King of England cannot, for in his dreams the King of England is precluded from being any king other than the one he is. His reality won't let him feel." (Pessoa 1982/2002: 438 - 439).

The idea suggested by this insightful passage seems to be the following one:[33] in order to successfully imagine being someone else, imaginers need to dispossess all their descriptive contents, i.e. their personal identity, the kind of individuals they are. Everything that characterizes them as a specific "someone" should be put aside in order to let them become "nothing", no one, i.e. a virtual self capable of surfing across an endless variety of imaginative perspectives.

[33] It goes without saying that mine is only one among various interpretations that could be given to this passage.

If the imaginer bears with him traces of his identity, like in the case of the King of England in Pessoa's passage, then it would be impossible for him to explore a different perspective from the one he is currently occupying, for instance, the perspective occupied by the King of Rome ("his reality won't let him feel").

The intuition underlying this passage seems to be that, given their *boundness* to the actual circumstances they are in and the defining characteristics they incorporate, our actual selves cannot imaginatively travel across different perspectives. Unless they dismiss any trace of the kind of selves they are in the actual world.

This idea has a straightforward intuitive appeal: if that was not the case, it would simply be impossible for us to explore perspectives other than our own. We would be, so to speak, irremediably confined to our small egocentric universe with no hope of looking outside of its self-referential experiential perimeter.

Interestingly enough, Pessoa's intuition is not confined to the context of literature but finds validation in some philosophical work on imagination. Specifically, defenders of the single code theory of imagination, such as Shaun Nichols (Nichols 2008), have argued that the self, or the first-person, interacts with imagination in a way that is at its core sympathetic with Pessoa's insight. The self, indeed, is taken by Nichols to be deployed in the imagination in an exceedingly impoverished way, i.e. deprived of almost any descriptive content that might characterize it.

The scope of this chapter consists in critically scrutinizing this thesis about the self and the imagination. I will try to provide arguments in order to counter Pessoa's insight and I will build my case against it by engaging with Nichols' analysis of the first-person and the imagination.

The upshot of my analysis will be that, in order to imagine being someone else – when by this we mean an imaginative exploration of someone else's cognitively and affectively centered perspective in a given situation – we need a much richer kind of self at play in the process. Indeed, if we want to grasp the first-person perspective of another individual, our *actual* thinking and feeling self needs to be engaged.

As we will see, Nichols' considerations pertain to a specific kind of imagination, which already happened to encounter as propositional or belief-like imagination. This latter kind of imagination, though, does not seem to capture the main sense of what it means to actually imagine being someone else.

What we usually mean by "imagining being someone else" is instead better captured by another kind of imagination, which I already identified with "empathetic imagination" or simply "empathy". As it should already be clear by now, empathy deals with representing another subject's perspective *from the inside.* This latter kind of process requires our actual self to be engaged in the imaginative representation in order to achieve an experientially rich appraisal of someone else's first-person perspective.

The main gain of the present chapter will allow me to highlight some major limits of empathic strategies. As I will try to prove throughout the remaining of the present chapter, what constitutes the condition of possibility of our imaginative exploration of intersubjectivity is also what constitutes its main limit. Given its ultimate boundness to our actual self, the empathetic imagination presents some insurmountable constraints that are grounded in the kind of individuals we are.

This constitutes the core of what Pessoa got wrong about the imagination – more specifically about imagining being someone else and, therefore, about empathy.

2. Indexicality and the "thin" psychological profile of the self

In his 2008 paper, Shaun Nichols addresses the issue of the interaction between the first-person and the imagination. The main theoretical aim of the paper is to provide arguments against the use of thought experiments for assessing questions linked to the nature of the self and personal identity. The reason for such skepticism is due to the peculiar way in which the indexical "I" interacts with the imagination. It is this latter aspect of the paper, rather than its argument against the use of thought experiments and imagination for assessing the nature of the self and personal identity that makes it relevant for our present purposes.

Nichols sets off his arguments through a critical assessment of some of Bernard Williams' most well-known thought experiments. In one of these experiments, the reader is asked to imagine the proposition "I am Napoleon", i.e. she is asked to imagine being an individual that does not share any of her physical and psychological characteristics. That seems to be a task that can be successfully accomplished: we can easily imagine that we are Napoleon.

However, beyond that apparent simplicity awaits a troublesome puzzle: if I imagine being deprived of all my physical and psychological traits, it would seem that I would cease to exist and that there would only be Napoleon (Nichols 2008; Recanati 2007; Williams 1973). How is it then possible that

I can so easily imagine that I am Napoleon if it is so manifestly impossible that *I* be him?[34] The answer is to be found in the peculiar way in which indexicals, and the indexical "I" in particular, interact with the imagination.

In the domain of linguistics and philosophy of language, the Kaplanian tradition conceives of the I as having a *content*, namely the specific person picked out by a token of the word and a *character,* namely a rule that identifies the content of any token of "I" with the speaker of the utterance (Nichols 2008: 221). It is clear, however, that this way of spelling out the indexical work of the "I" in terms of a content/character couple is not adequate to account for its work in the psychological domain, i.e. in the context of conscious thought and behavior. The way in which the I operates in the context of cognition is indeed far from being fully captured by its semantic function in the context of language.

Nichols' strategy to identify the role of the I in the psychological realm is to conceive of it as a sort of mental vehicle that "marks" those thoughts that could be expressed by sentences that include words such as "I" or "me" (221). These mental symbols, called "I-concepts", serve the purpose of picking out the self in a way that is particularly significant in the context of e.g. action production.

To paraphrase Nichols' example: suppose that I have a belief that MI is about to be attacked by a bear. My disposition to act would greatly depend on the fact that I do or do not believe that *I* am MI. If the thought I am having is marked by the I-concept, viz. if it is linked in a self-referential way to me as the *subject* who is having that thought, it is very likely that the subsequent behavior would be significantly determined by it, e.g. *I* would start running or planning a strategy to defend *myself* from the imminent threat.

These considerations are meant to be fairly uncontroversial and introduce the main, theoretically more robust claim of the article, namely that the psychological profile of the self picked out by the I-concepts is remarkably thin. I-concepts can indeed be at work even in cases when almost every feature pertaining to the subject having them are utterly unbeknownst to him:

> A person can wake up in darkness with total amnesia yet think *I have a headache*. The I-concept is functioning normally here, but the agent has no distinctive descriptive content associated with *I.* (523)

---

[34] Nichols goes on to analyse the other two very famous examples given by Williams, but the Napoleon-case is already sufficient to grasp his argument.

The subject uttering the sentence in the above example has no idea about who they are and where they are, but they are still in the position to have properly functioning I-thoughts, or, so to speak, to acknowledge some kind of psychological "ownership" on the experiences they are undergoing.

The exceedingly thin psychological profile of the self does not imply that it is completely devoid of content, though. This would, indeed, make it a purest and empty placeholder for something utterly undetermined. As Nichols carefully notices, the I-concept seems to carry *some* minimal categorical information about the subject: the referent picked out by the I should at least be generally qualified as an *agent*. That would also explain why sentences such as *I am a gas molecule* sound somehow bizarre or surprising.[35]

3. The self in the context of imagination

The crucial question now is to assess how the self interacts with the imagination. Three features of the imagination are needed in order for Nichols' arguments to achieve their goal:

i) Imagination has to be conceived in representational terms. In other words, having an imagining of *p* means having an *imaginative representation* with the content *p*.
ii) Imaginings and beliefs are to be distinguished not by their contents, but by their functional roles. I can indeed hold at the same time a belief and an imaginative representation with the exact same content, yet still be able to identify them as two clearly distinguishable mental states. I can indeed believe that the cup in front of me is empty and at the same time *imagine* it is empty after having partaken in a pretend-game in which a make-believedly full cup of tea became empty after someone spilled its content on the table (Leslie 1994).
iii) Imagination entails, to a large extent, the operation of the same mental mechanisms that are deployed in belief states. Imagination and beliefs are therefore treated similarly by those mechanisms:

> For instance, if a mechanism takes imagination representations as input, the single code hypothesis maintains that if that mechanism is activated by the occurrent belief that *p*, it will also be activated by the occurrent imagination representation that *p*. (Nichols 2008: 525)

It is precisely iii) that plays the most significant role in the overall argumentation of the paper. Indeed, as we would expect by the single code hypothesis, if it is true that "I believe that *p*" deploys roughly

[35] Sure, we can easily imagine a fictional story starting with *I am a gas molecule*. At the same time though, it would be obvious that the gas molecule of the tale would be immediately endowed with some minimal agential properties such as the capacity to express a sentence about itself.

the same cognitive mechanisms as "I imagine that *p*", we would expect that the indexical "I" in the imagination case behaves in precisely the same way as in the belief case, i.e. in its greatly impoverished way. For instance, the *belief* "I have a headache" behaves similarly to the *imagining* "I have a headache". The self in the imagination is presented with the same minimal representational format displayed in the cases of beliefs: it bears no actual descriptive contents about the subject who is actually having such mental states.

This allows us to explain why contradictory imaginings can be plainly produced without us being immediately aware of any relevant epistemic obstacle impeding them. I can imagine that *I am Napoleon* simply because the I is deployed in its highly minimalistic form. This is at the basis of "the supreme flexibility of the "I" that allows for the logically problematic representation that "I exist in this location with completely different psychological properties" (Nichols 2008: 527), such as in the case of Napoleon.

The upshot of this argumentation is that we should be extremely skeptical about the epistemic fruitfulness of deploying the imagination in order to assess the nature of the self: the impoverished manner in which the I interacts with imagination has indeed epistemically disastrous consequences such as letting us imagine that our self would persist even if we were to deprive it of all its descriptive and historic-causal features that define it – like in the case of Napoleon.

The extreme flexibility of the indexical "I" usually does not create any major trouble in the case of beliefs because it is always clear who is the subject of the belief-state, i.e. the subject holding such belief. In the case of imagination instead, and specifically in the case of imagining being someone else, an ambiguity is created that problematically convolutes the indexical work of the I-concept.

It is indeed less clear which subject is the I-concept actually picking out: who's the I in the Napoleon-imagining?

4. A "thicker" version of the self

As already announced in the previous section, it is not the case against the deployment of thought experiments to assess the nature of the I and personal identity that interests me here but the claim that grounds it, namely the minimalistic psychological profile of the self and its peculiar way of interacting with the imagination.

I substantially agree with the arguments presented by Nichols: the self-referential role of the I can be accomplished through a minimally determined psychological vehicle whose role is to merely self-

ascribe the proposition imagined. At the same time though, it is undeniable that there are other ways in which the self can interact with the imagination. In these other cases, a psychologically much richer profile of the self needs to be recruited for the subject to carry out the relevant imaginative enterprise.

Before moving to address this point though, it might be useful to clarify some features of the imaginative process we have been dealing with so far. As we saw by engaging with Nichols' argumentations, there is a peculiar way in which the self figures in imaginative representations like those performed when I am asked to imagine that I am Napoleon.

Now, in order to make the overarching picture clearer, we need to be aware of two distinct kinds of self that are simultaneously involved in the imaginative process:

i) the self of the agent doing the imagining, i.e. the cognitive agent generating the imaginative representation. This self is only *externally* related to the imagining as the cognitive producer of it;
ii) the self that figures *within* the imaginative representation, i.e. the self who is *internally* related to what is imagined as constitutes the subject the imagined proposition is centered on.

To distinguish these two selves I will call them (i) the *actual self* and (ii) the *imagined self*, where the former is the cognitive agent and the latter the subjective center of the imagined proposition.

In the cases described by Williams and discussed by Nichols, the imagined self is supposed to pick out the actual self and make it the center of the imagined proposition (i.e. the imagined self), while at the same time identifying it with another individual self from which it is metaphysically distinguished (i.e. Napoleon).

As we saw, Nichols' take on this puzzle consists in pointing out that the imagined self not only bears none of the distinguishing features constituting the actual self but is also so minimally characterized that it can be imaginatively identified with other individual selves without that generating an intuition of manifest impossibility. The imagined self can, therefore, travel across all sorts of imagined scenarios as a Cartesian ghost, capable of swiftly eluding the otherwise inflexible restrictions of the laws of metaphysics.

However, there is at least another kind of imaginative process where the imagined self interacts with the actual self in a much more robust way. The cases discussed by Nichols, indeed, are captured by a specific form of imagination, which – as we also saw in the *Appendix* – has been called by scholars in the field belief-like or propositional imagination.

The case of propositional imagination represents, though, only one possible form that the imagination (broadly conceived) can take. Therefore, narrowing down the scope of what it means to imagine being someone else to the mere case of belief-like imagining being them would be a rather limited way to address the issue. Furthermore, if we read more carefully Williams' passages, we see that the kind of imagination he alludes to cannot be plainly reduced to an instance of propositional imagination but seems to be rather concerned with something more closely related to modeling the first-personal perspective of another individual, in our specific case, Napoleon:

> "Images of myself being Napoleon can scarcely merely be images of the physical figure of Napoleon ... They will rather be imagines of, for instance, the desolation of Austerlitz as viewed by me vaguely aware of my short stature and my cockaded hat, my hand in my tunic." (Williams 1973: 43)

Williams seems to allude to the use of imagery (e.g. the desolation of Austerlitz) and the representation of the subject's cognitive and affective dimensions (e.g. the awareness of the short stature, the cockaded hat, the hand in the tunic).

As it seems, when trying to imagine being someone else, what we are usually and foremost interested in is certainly not adequately captured by mere belief-like imagination. When we imagine being someone else, usually we do it for the purpose of mutual understanding. Even if propositional imagination can bring some significant contribution to our processes of mutual understanding as stated by Nichols himself (Nichols 2019) and as I also noticed in the *Appendix*, it still does not seem to capture the scope of the most common occurrences of imagining being someone else. When trying to represent being someone else, we imaginatively transpose ourselves in her shoes in order to achieve a representation of her first-personal perspective and gain insight into her specific "take" on a particular situation she happens to experience.

The work of modelling someone else's first-person perspective is accomplished by what we can call empathic imagination (or simply, empathy) as opposed to belief-like imagination. As insightfully stated by Richard Moran:

> [Propositional imagination] only needs to be provided with the proposition in order to reason from the assumption of its truth. By contrast, imagination [the kind of imagination that is much closer to what we have called empathetic imagination] involves something more like a point of view, a total perspective on the situation, rather than just the truth of a specifiable proposition. And imagining along these lines involves something more like genuine rehearsal, "trying on" the point of view, trying to determine what it is like to inhabit it. It is something I may not be able to do if my heart is not in it. (Moran 1994)

Far from being a mere instance of counterfactual reasoning, where all we need is the right proposition, empathy requires us to bring ourselves in the process in order to achieve a first-personal appreciation of it. With a formula, we could say that empathetic imagination is not merely a matter of content, i.e. *what* is imagined, but a matter of *how* it is imagined, i.e. how we imaginatively engage with the contents represented (Camp 2017).[36]

Empathically imagining being someone else engages the self in a remarkably different way that is largely incompatible with the "exceedingly poor" psychological profile of the self envisioned by Nichols' theory (and suggested by Pessoa's remarks on imagination). With this kind of imagination, our selves are engaged in a way that does not reduce in any way to the role of a minimally qualified referent of a mental indexical.

In this kind of process, the set of psychological resources involved are likely to be considerably different from those required to merely propositionally imagine that a certain proposition with a certain content is true. As stated by Kendall Walton (2015:38), when we imagine – in the sense of what we called empathetic imagination:

> we imagine *doing* things, *experiencing* things, *feeling* in certain ways. We bring much of our actual selves, our real-life beliefs and attitudes and personalities, to our imaginative experiences.

In order to model a subject's perspective in a given situation, we have to deploy cognitive processes that are somehow similar to the ones that would be deployed were we ourselves, as possible subjects of similar experiences, to occupy the target's vantage point under her specific circumstances. Stated in our words, in these contexts, the actual self seeps into the imagined scenario and substantially contributes to model the imagined self by providing it with a considerable part of the required psychological apparatus.

4.1. More on the influence of our actual self on our imagined self

The point raised at the end of the previous section is a pretty delicate one. It has not only represented a matter of controversy in debates about perspective-taking, but also constituted a reason for widespread confusion about what it means to imagine being someone else.

---

[36] This kind of imagination has also been called "dramatic rehearsal" (Moran 1994).

In both the psychological and philosophical literature, one indeed finds plenty of work concerned with assessing the role of the actual self in the process of imagining being someone else. As we already anticipated in the first chapter of the present work, a way in which the discussion is usually set consists in individuating two closely related processes, which are nevertheless discernable and show remarkable differences in terms of both their epistemic outcomes and behavioral, psychological and neurological profiles.

As anticipated in *Chapter I*, there are at least two possible ways in which we can conceive of the process of imagining being someone else: I can imagine *myself* in somebody else's shoes or I can imagine *being* somebody else in a specific circumstance. Following Goldie's terminology, the first kind of process can be called "in-his-shoes perspective-taking" and the latter "empathetic perspective-taking" (Goldie 2011). The difference between the two imaginative projects can be spelled out by referring to the role played in the imaginative process by what I called the actual self, i.e. the self who is actually doing the imagining. When we perspective-take along the lines of "in-his-shoes perspective-taking", we merely imagine what *we ourselves* would do if faced with the circumstances the other person is currently facing, i.e. we import our actual self as such within our imaginative representation. In this case, the imagined self substantially maintains the same psychological profile of our actual self and is taken as a proxy to how we would feel or behave if we were to experience those imagined circumstances. When we perspective-take along the lines of what Goldie calls "empathetic perspective-taking", we do not import the distinctive traits of our actual selves within the imaginative representation but rather try to imagine how *the other person* – with her distinctive psychological features – would feel or behave in a specific situation. Hence, with empathetic perspective-taking, the imagined self does not bear traces of our actual self and is modeled alongside the specific psychological profile of our target.

This distinction is echoed in some relevant empirical work in the field of social psychology and neuroscience: scientists in these fields usually distinguish between imagine-self and imagine-other perspective taking.[37] In a series of studies run by Batson and colleagues (Batson et al. 1997) participants were asked to perspective-take by means of either imagine-self or imagine-other processes. It turned out that the two processes lead to different psychological and behavioral outcomes. In the imagine-self condition, people showed a greater focus on themselves: they reported higher levels of personal distress and tended to avoid helping another person in need if it was easy to

[37] See Coplan 2011 and Maibom 2020 for a philosophical overview of such studies.

escape the situation in which their help was needed. In the imagine-other condition, they instead reported a lesser degree of personal distress, showed more concern for others and were more willing to help someone in need even when it would have been easy for them to avoid doing so (Batson 2011). This distinction has also been mapped in a series of neuroscientific experiments using the fMRI technique to detect the brain activity related to the different perspective-taking tasks: different brain areas are activated when performing respectively imagine-self and imagine-other tasks (Decety and Sommerville 2003; Ruby and Decety 2001, 2004). To put it succinctly: it seems that Goldie's distinction is not only philosophically legit but also empirically warranted.

We can now try to re-assess the considerations I have been offering so far about the role of the self in imagining being someone else. In light of Goldie's distinction and its empirical validation, one might indeed be tempted to conclude that what I have been referring to as the actual self is the one that figures only when we are performing the tasks of in-his-shoes perspective-taking or imagine-self, whereas empathetic perspective-taking or imagine-other processes guarantee that the imagined self be psychologically thin enough to take up an utterly different identity from the actual self. I believe, though, there are good reasons to reject this interpretation and to be rather skeptical of a clear-cut distinction between imagine-self and imagine-other processes. But let us proceed in order.

As insightfully noted by Ylwa Wirling (Wirling 2014), when trying to evaluate an account of what it is to imagine being someone else, there are at least three questions to be asked:

1) What role does the imaginer's self play in the imagining? Or, in our terms, what role does the actual self play?
2) What is the imagined relationship between the actual self and the imagined self?
3) Is the distinction between in-his-shoes perspective taking and empathetic perspective-taking maintained?

We can rather easily track the answers to these questions that are implicitly embedded in Nichols' account. He claims that the imagined self bears no trace of the actual self (1), that there is no significant relationship between the actual self and the imagined self, since the latter works merely as a minimally defined center of subjectivity that can be used to take up a variety of imaginative identities (e.g. Napoleon's) (2), and that finally, his account does not allow for any significant distinction between imagine-self and imagine-other processes, since the self figuring in the imagining is so minimally defined that can be easily interchanged between *my* specific fist-person perspective and *anyone else's* first-person perspective (3).

My present account instead, as previously anticipated in different terms, holds that:

1) The actual self substantially influences not only the process but also the content of what is imagined.
2) The imagined self is *bound* to the actual self.
3) The boundaries between imagine-self and imagine-other are more blurred than one might think. Imagine-other tasks cannot rely on the capacity to deprive our imagined self of all the properties that define our actual self.

The overarching principle governing the above three answers is the following: when trying to imagine being someone else, even if by this we intend to perform an act of empathetic perspective-taking (or imagine-self), we inevitably feed up our imaginings with properties and traits that are our own, i.e. belong to our actual self. Let me try to be more articulate about it.

It seems natural to relate the above considerations to the longstanding debate about the simulation theory, which has constituted for decades a highly debated aspect of perspective-taking and mindreading in general (A. I. Goldman 2006; Spaulding 2016). Far from constituting a homogeneous set of claims, simulation theory has to be conceived of as a rather diversified approach to the issue of mutual understanding.

Notwithstanding a remarkable diversity in the multifarious ecosystem of simulation theory, the presence of some shared fundamental assumptions behind the different approaches is undeniable. Among these underlying common features,[38] one that is particularly relevant for present purposes is the claim that, in order to understand other people, we deploy roughly the same cognitive resources that would be at play if we were to actually occupy their position in analogous circumstances. The cognitive resources recruited by the simulator, though, are said to be deployed *offline*, i.e. largely decoupled from the spontaneous causal impact they would normally display were she to actually hold such mental states.

If we accept the underlying general assumption according to which, in order to simulate someone else's mind, we resort to *our own* mind as a model, it seems to follow that we would still find a substantial underlying coherence with our actual psychological profile among the set of possible simulative processes we engage with. This means that, although we can feed our imagination with

[38] Whose list, of course, I neither have space nor interest to spell out exhaustively.

pretend-states that would allow for the minimal imaginative flexibility required to project ourselves in possible, non-actual, or utterly fictional scenarios, we would nevertheless be prey of some form of psychological rigidity that can be ascribable to the actual individuals we are, i.e. to our actual self.

Prima facie, simulation theory seems to provide us with a nice account of what it means to imagine being someone else and that seems to capture the essence of the idea I am after. However, even if I certainly believe that simulation theory provides us with useful conceptual tools to adequately frame some of the central features of empathy, appealing to the essential principles of simulation theory I just described above without further qualifications will inevitably generate some confusion and prevent us from properly grasping the issue at stake. One should indeed distinguish between the *process* that allows us to imagine being someone else and the *content* of what we imagine when trying to take someone else's perspective. Simulation theory allows us only to claim that the *processes* or mechanisms involved in the imagining being someone else are the same we would deploy, were we to actually occupy out target's perspective. In other words, the mechanisms employed in imagining being another are also those employed in actually being ourselves. But here is precisely where the risk of confusion lurks: these mechanisms, as described by simulation theorists, are just common bits of human cognitive architecture, they are not distinctively *ours* or our *self*. They are assumed to be shared among individuals and therefore to be at the basis of the interchangeability of points of view, rather than an element for the individuation of personal psychological profiles. Thus, the conceptual tools we borrowed from simulation theory seem to be poorly placed to make a case for a "thicker" self in the imaginative process, if we do not further qualify them.

However, we can look a bit deeper on the process of imagining being someone else and discover that the way in which we generate the imagining (i.e. the process) is somehow always related to our actual self. Indeed, our empathetically mediated understanding of the other person will dramatically depend on the kinds of properties we ascribe to her and these will, in turn, depend on the way in which we *scaffold* the imagining. In this sense the *contents* that figure in the imagining are determined by how we run the imaginative *process*. That is, even if we grant that the mechanisms involved in simulating are common bits of human architecture and, therefore, do not bear the mark of the actual self, the way in which they are employed and the contexts in which they figure are essentially related to our actual self. The psychological properties that we take an individual to instantiate in a given situation are, indeed, likely to be in continuity with the properties we would display in similar circumstances. To use our favored example, I usually respond with amusement to a dog wagging its tale. Now, if I were to imagine someone else undergoing that experience, *ceteris paribus*, I would ascribe to them the same feeling of amusement I would experience.

One might argue that this would merely be just in-his-shoes imagining. It does not capture cases in which we try to re-orient our egocentric map alongside our target's psychology. The claim I want to defend here, though, is ambitious enough to maintain that our egocentric map can surely be reoriented but not beyond restraints and that these restraints are actually determined by our actual self. This point becomes more evident if we actually come back to the imagine-self/imagine-other distinction. As said before, even though this distinction is philosophically solid and empirically warranted, we have good reasons to doubt that it is something that clearly maps on the way we actually perspective-take. Indeed, when we are trying to imagine being someone else, rather than imagining how we would act in a non-actual situation, we are likely to mediate between psychological features and attitudes that are partly ours and partly taken to be specifically of the other. In a recent publication, Heidi Maibom (Maibom 2018) made this point very clearly:

> The imagine-other vs. imagine-self distinction between forms of perspective taking operates with a relatively rigid idea of separation between individuals. Either I imagine *you*, and just *you*, or I imagine *me*, and just *me*. But doing only one or the other rather defeats the purpose of perspective taking in the first place. For I am supposed to blend you and me by applying my own psychological process to your situation. I am not supposed to simply project myself into your situation. (Maibom 2018: 77)

This sounds all the more appropriate if we consider that it is rarely is the case that we have available the full range of our target's mental states and attitudes in a given situation. Most of the time we have only scarce bits of information about the other person's and the situation they may find themselves in, so that we have to "fill in the gaps" and this of course will depend on our actual self's psychology, its attitudes, values, thinking routines, etc. Most of the times, even if individuals have access to the same kind of information, they interpret and frame it differently depending on a remarkably wide variety of variables such as their underlying motives, beliefs, values, political views, perceived social identity, etc.

This is perfectly illustrated by a recent tragic event which was a major source of controversy within the American public opinion and that is adduced by Shannon Spaulding at the beginning of her 2017 paper about individual differences in mindreading (Shannon 2018). The event involved a police officer who, on July 17, 2014 in Staten Island, killed an African American man who had been stopped for allegedly selling loose cigarettes after holding him for several minutes in the chockehold position. The man repeated eleven times "I can't breathe" while the police officer was holding him. Putting aside the immediate political and social implications of such a dismaying incident of violence, Spaulding is mostly interested in the interpretations that people have given of this event. The entire unfolding of the interaction between the man and the police officer was entirely recorded on video.

For this reasons, one would reasonably expect that there would not be much space for disagreement, since everybody has access to a tape neutrally capturing how things really went: people could evaluate for themselves the details of what really happened. Unfortunately, precisely the opposite occurred: the public opinion was polarized among two parties, those who saw the man's behavior as dangerously hostile and therefore interpreted the police officer's behavior as fully legitimate and those who interpreted it as an unacceptable and abusive display of police violence.

People were watching the same video, they had in principle access to the same kind of information and yet they could not disagree more. An exemplary demonstration of such disagreement took place at the radio program This American Life (episode 548), where two women were asked to watch the video with the intent of finding an agreement about how best to interpret it.[39] One was an African American producer for This American Life with past working experience in a correction facility while the other person, a friend of hers, was a white New York City police officer. After several hours of discussion, the two women could not reach any kind of substantial agreement on how to interpret the event. Clearly, the two women were framing the event in two different, incompatible ways. That is Spaulding's suggestion on how to make sense of such an episode:

> Though it is difficult to say with any precision exactly what features of a situation two particular individuals are taking as input for mindreading, we can take the disagreement reported on This American Life as a schematic example of how such disagreements arise. The producer for This American Life likely identifies with Garner as part of her racial in-group, whereas the police officer likely identifies with the other police offers as part of her in-group. […] [T]his subjective sense of perceived similarity influences how one sees and interprets a target's behavior. Furthermore, the past experiences and narratives familiar to the producer and the police officer likely differ, so they will notice different features of the situation and interpret those features in light of narratives familiar to them. Though the producer and police officer are looking at the same interaction, they are attending to different aspects of the situation and interpreting those aspects differently. (Spaulding 2018: 4022)

The two women, in virtue of a wide variety of personal variables such as their differing personal narratives, moral concerns, and perceived social identity were processing the same information in a way that did not leave space for agreement. They were filling in the gaps about the man's and police officer's mental states based on their two different ways of reading through the situation they were witnessing. In virtue of their diverging approaches to the same event, they were selecting and

---

[39] Here is the link to the episode: http://www.thisamericanlife.org/radio-archives/episode/548/ cops-see-it-differently-part-two?act=0\#play.

emphasizing particular elements of the interactions that were likely to lead to diverging conclusions on how to best interpret it.

There is a lot to unpack in the example provided by Spaulding and part of these unanswered questions will be tackled throughout the remaining of this work. At the moment, though, suffice it to focus on the intuitive grip that this story has on the issue that interests us the most at this point. The way in which we frame our understanding depends on personal variables that play out differently in mindreading processes. Stated otherwise, our actual self plays a pivotal role in framing and guiding the process of understanding. Even if these considerations are meant to apply to the whole range of mindreading capabilities in general rather than a specific mindreading strategy (e.g. empathetic imagination), there is no principled reason to exclude the fact that such variables do also influence the way in which we imagine being someone else. Here as well, indeed, we "fill in the gaps" and selectively deploy information to build our imaginative representation of another's perspective.

As I will try to show in the next section, the problem can be even more radical than that. Indeed, there might be cases where we have access to a wealth of information about the other's situation and mental states, i.e. we can be provided with all the relevant details about the event we are trying to imagine. And yet, we can still fail to empathetically imagine these perspectives because they are in conflict with our actual self. If a perspective requires that we imaginatively "migrate" too far away from our self, our imaginative projects shrink and, often, collapse. As we will see, the constraints posed by our actual self requires us to "come to terms" with whom we are in order to achieve an empathic understanding of the others, i.e. we have to bring our selves to imaginatively inhabit the perspective of the other person and this. As we will see, this is not always an available option.

Let us pause this for a moment before moving on. My arguments should not be read as irreversibly downplaying the imaginative flexibility of imagining minds. Even though there certainly are limits to how far can someone project oneself into an alien perspective, the domain of experiential states we can imagine can be impressively rich, depending on the kind and amount of information we gather about our target's internal and external circumstances. However, we have to acknowledge that the imagined self engaged by the empathetic imagination is far from having the minimalistic profile that Nichols described in his paper. Our actual, psychologically "thicker" self does play a major role in imagining being someone else: it constitutes the fundamental psychological engine of such processes and provides our representations with the "raw material" to model another's first-personal perspective.

In order to imaginatively occupy someone else's perspective, we have to embed our imagined self in the target's situation and try to work out what it feels like to be in their shoes. For this task to be accomplished, our actual self needs to be directly engaged to some extent in the imaginative representation. The way in which we scaffold and fuel our imaginative representations of alternative perspectives are parasitically dependent on the kind of individuals we are. The fact that we have a self that is much more richly characterized than a mere psychological indexical, is what influences and shapes the way in which we imagine being someone else.

This may sound a bit too abstract. But we can actually build an intuitively richer case by providing concrete examples that illustrate this point more efficiently, i.e. cases where the obtrusion of the actual self in the imagined process seems to recurrently happen with striking evidence in various contexts. Also, turning to some relevant work in social psychology will provide our argument with a persuasive empirical basis that goes beyond pure armchair speculations.

This is what will keep us busy in the remaining sections of this chapter.

4.2. The actual self in empathy: the case of imaginative resistance

The concrete example I have in mind is the so-called phenomenon of imaginative resistance. The occurrence of imaginative resistance can properly be explained by referring to a psychologically richer kind of self being directly engaged by the empathic imaginative process.

A considerable amount of work has been produced on the philosophical puzzle of imaginative resistance and scholars have used a variety of strategies to address it (or explain it away). For these reasons, entering an already quite crowded area of debate is not my aim here. Thus, rather than problematizing or trying to offer a philosophically appealing solution to the puzzle of imaginative resistance, I am going to acknowledge its occurrence and prove that the best way to explain it is through the engagement of a much more robustly characterized self in processes of empathetic imagination.

The phenomenon of imaginative resistance is puzzling because it reveals an irregularity in the functioning of imagination that cannot be easily settled. Here is the puzzle: the flexibility of our imagination seems to allow for all sorts of oddities to be imagined. We usually do not show the slightest bafflement when asked to imagine a piece of wood casting magic spells, caterpillars smoking hookahs, and that the devil and his lackeys are plunging Moscow into chaos with dark magic. Imagination grants us the ability to effortlessly surf across counterfactual scenarios, fictional worlds, alternative perspectives. However, when it comes to imagining morally deviant cases, the

representational tolerance of imagination startles. When we are asked to imagine, say, that "murder is morally praiseworthy" or that "torturing babies is a pleasing step towards a life well lived" our imagination limps: it does not allow us to explore such scenarios, i.e. it resists them. How come?[40]

As a general premise, it is necessary to notice how these cases of resistance do not seem to emerge in instances of mere belief-like imagination (Leeuwen 2016). Indeed, I am going to grant that imaginative resistance occurs only when we are asked to imaginatively represent certain states of affairs in an experientially richer manner, i.e. when asked to first-personally represent that certain morally deviant states of affairs obtain in a world that we imaginatively inhabit. For instance, if asked to imagine that "unmotivated murder is morally praiseworthy" we do not simply have to suppose that this proposition is true, but we have to represent a world involving ourselves as experiencing subjects, in which the state of affairs encapsulated by the proposition holds true *for us*. To stick with the terminology we deployed so far, imaginative resistance is likely to occur in the case of empathetic imagination.

Several explanations have been proposed in the specialized literature to account for the puzzle of imaginative resistance. In a recent overview of the literature on this topic, Tamar Gendler and Shen-yi Liao (Gendler and Liao 2016) grouped different solutions to the problem of imaginative resistance into three different sets: the wontian, cantian, and eliminativist theories. The labels assigned to the groups speak for the kind of explanatory strategy used by its tenants. Wontians claim that people experiencing imaginative resistance simply do not *want* to imaginatively engage with the content of the propositions they are asked to imagine. Cantians believe that people simply *cannot* imagine such morally deviant scenarios, i.e. their representation capabilities fall short from grasping how certain states of affairs could actually obtain. The eliminativist strategy instead consists in deflating the puzzle of imaginative resistance by showing how it can be explained away as a mere by-product of fake philosophical constructions that have no actual counterparts in real life. Eliminativists believe that the sentences that are used as compelling examples to evoke imaginative resistance are cleverly de-contextualized by the authors in order to artificially create the effect they then go on to explain. Outside the limited and exotic perimeter of philosophy papers, the eliminativist holds, such

[40] Here, I am going to focus on cases of imaginative resistance that are limited to morally deviant cases. Several scholars, though, believe that the phenomenon has a much broader scope that extends beyond the mere domain of morality (e.g. Weatherson 2004). Even though I tend to agree with these authors, I do not have to take a position in this particular context. Indeed, if we grant that imaginative resistance occurs *at least* in cases where we are asked to imagine morally deviant perspectives, then my point about the major obtrusion of the actual self is already proven without having to commit to an expanded account of imaginative resistance.

propositions almost never appear out of context and are made fully intelligible by the narrative – and, more generally, informational – background that surrounds them (Gendler and Liao 2016).

From this brief survey, it appears that the solutions that could best conform to our argumentative aims are the wontian and the cantian. As for what concerns the wontian theory, the reason is straightforward: if we do not imagine endorsing certain kinds of beliefs or embracing what we hold as repugnant moral principles because we do not *want* to engage with them, not even through our "safer" imaginative avatars, then there is a clear major interference of the actual self on the imagined self. The values and moral commitments of the former are imported and incorporated by the latter so that it cannot enter a perspective that it is taken to be as morally repellent.

A certain interpretation of the cantian theory also proves a major involvement of the actual self in the imaginative representation. Indeed, if we cannot imagine holding morally deviant beliefs because we simply cannot attune to perspectives that are utterly incompatible with our worldview, then again, the actual self is exerting major constrains on the imagined self. The actual self would export much of its actual psychological features, attitudes and proclivities to the imagined self, thereby proving the considerably thicker nature of the latter in the imaginative process.

Now, we do not need to fully embrace either a wontian theory or a cantian theory in order to prove our point. It is sufficient to acknowledge that some central occurrences of imaginative resistance can be captured by either a wontian or a cantian solution since both of them would prove the presence of a more richly characterized imagined self.

The eliminativist explanation, instead, does not allow us to deploy the case of imaginative resistance as an example because, as we said, it simply denies its real occurrence. Moreover, the eliminativist could even play a role in reinforcing Nichols' point (and Pessoa's insight): given its impoverished nature, our imagined self can be filled in with whatever psychological identity we like, no matter how distant from our actual self it is. The constrained or unconstrained flexibility of the self in imagination would simply be a contingent issue of information availability.

Most scholars in the specialized literature are mainly interested in explaining occurrences of imaginative resistance in the context of literature but there are very good reasons to think that this phenomenon is not restricted to the mere fictional world (Stueber 2011). Indeed, if we look at real-life instances of being confronted with seemingly incompatible perspectives, the obtrusion of the actual self in the imagined representation – not to mention the lack of the relevant information about the alien perspective – tends to be the rule rather than the exception.

In many circumstances we are in fact faced with incompatible and apparently incomprehensible perspectives and stuck with questions like “how on earth could someone believe such absurd thing?” or “how could they do something like that?”.

This peculiar configuration of imaginative resistance has been dubbed as the “phenomenological puzzle” (Gendler and Liao 2016; Van Leeuwen 2016), i.e. the experience of jarring, shock or confusion when confronted with such radically different perspectives.

To offer a real-life example, on the occasion of the terrible terror attacks in France on 13 November 2015, the General Secretary of the “Organisation internationale de la Francophonie” Michaëlle Jean declared:

> I am deeply *shocked* and *dismayed* by the cowardly and odious attacks that cause suffering and trauma. I condemn in the strongest terms these *immoral* and *senseless* attacks intended to spread terror across the world. These unacceptable and *unjustifiable* attacks that occur in the aftermath of those perpetrated in Beirut remind us of the need to resist together to the scourge that is terrorism. [emphasis added][41]

This declaration can be interpreted as a concrete occurrence of imaginative resistance. Jean found herself “shocked” and “dismayed” and the actions perpetrated by the terrorists seemed to be so atrociously “immoral” that they appeared “senseless” and “unjustifiable”.

A die-hard eliminativist would still be in the position to claim that, if provided with the relevant background information about, say, the peculiar life history of the terrorists, their path toward Islamic radicalization and maybe some insight into the psychology of the radicalization, Jean – or anyone else – would not experience imaginative resistance anymore. This could turn out to be a fair move – even though the eliminativist would have to work to find evidence to support the claim. In any case, I am willing to grant that the lack of the relevant informational background could surely amplify the bafflement one experiences in such cases. Still, merely acknowledging the occurrence of imaginative resistance deprives the eliminativist solution of much of its argumentative power: it proves that imaginative resistance is not the artificial by-product of epistemically unfruitful philosophical elucubrations, but rather a real-life phenomenon.

To provide an even more convincing argument against the die-hard eliminativist though, we can look at cases where, even if the subject is provided with reasonably enough information about the other person’s perspective, we can still experience imaginative resistance. Even if, in such occasions, we

[41] https://www.haitilibre.com/en/news-15763-haiti-flash-paris-attacks-messages-of-sympathy.html

succeed in identifying the relevant mental attitudes that ground another person's specific take on the world, we might still fail at getting a grip on *how* it could be so, i.e. how such attitudes could be held from a first-personal point of view and motivate certain conducts.

The clash between atheists and religious people turns out to be useful in this context. During a conference, Richard Dawkins made this statement:

> I find it extremely hard to *imagine how* any creationist that actually bothered to listen to that [i.e. to a sound evolutionary story], could possibly doubt the fact of evolution.

Dawkins claimed that he finds it impossible to imagine how a person provided with the right information about evolution could still doubt its truth and he seems relentlessly puzzled by the fact that such individuals do exist.[42] Later on, in the same interview, he defined minds like these – minds that are not persuaded by evolutionary explanations to give up their creationist beliefs – as "a disgrace to the human species".[43]

Dawkins' case seems to be more naturally captured by a cantian rather than by a wontian reading. Indeed, the problem with Dawkins is not that he does not want to understand the worldview of a tenacious believer confronted with the theory of evolution, but that he literally cannot imagine *how* someone could still believe in the creationist story after having learnt about evolution.[44]

In any case, from these real-life cases, we can safely infer that it does not really matter whether people like Dawkins or Jean *cannot* or *do not want* to imagine certain perspectives, their imaginative resistance can still be explained by the conspicuous obtrusion of their actual self in the imagined perspective: how else could we explain our reticence or inability to try on perspectives that dramatically clash with our own worldview?

Surely there is an upstream explanation to be given as to why people refuse to take on mental attitudes that motivated behaviors that are dramatically incompatible with their worldview.[45] For our purpose

---

[42] Specifically, these considerations were referred to the geologist Kurt Wise.

[43] For the video where Dawkins made such statements see:
https://www.youtube.com/watch?v=TjxZ6MrBl9E&feature=emb_logo.

[44] Surely a wontian might still want to claim that Dawkins does not try "hard enough" to get the creationist viewpoint, proving that his imaginative resistance is a matter of will rather than a matter of representaitional capabilities. However, I still believe that the most charitable interpretation of Dawkins' puzzlement is more naturally provided by a cantian reading of imaginative resistance. Indeed, he has reasonably enough information about creationists' beliefs to "try on" their point of view.

[45] For an overview of upstream explanations of imaginative resistance, check again Liao and Gendler 2016.

though, I believe it is enough to merely acknowledge the occurrence of this phenomenon. This provides some intuitive evidence supporting the case that our imagined selves show a remarkable rigidity in virtue of the psychological import of some relevant features of our actual self into the imaginative process.

Since first-personal imagination requires different cognitive and affective resources than merely supposing that certain propositional contents are true and given that these resources can be retrieved by engaging our actual self in the imagined scenario, it follows that the imagined self will inherit at least some significant features of the actual self.

If what we said is on the right track, then it is obvious that reluctance to imagine certain perspectives would arise. The imagined self is an imaginary re-styled version of the actual self but, at its core, the former remains bound to the latter.

4.3. Some empirical support

Before wrapping up what has been said so far, it might be useful to report some additional support to the arguments we presented so far. Such support comes from work done in cognitive science on the nature of imagination and perspective-taking, specifically on the ways in which our imagination interacts with other cognitive and affective systems in our brain.

There is a large body of evidence showing a robust parallelism between imaginative states and their non-imaginative counterparts. When we are asked to imagine non-actual states of affairs, interesting similarities occur between the merely imagined states and the actual states they aim to reproduce. For instance, similar brain areas are engaged both when we are in a certain perceptual state and when we are asked to *imagine* being in a certain perceptual state. Imagining *seeing* a certain scene activates brain areas that are linked to online processes of vision. The same logic seems to hold true not only for other sense modalities but also for our motor systems: when imagining performing a certain motor task, areas of the motor cortex are significantly engaged.[46]

On the one hand, this evidence could be invoked to further support the idea of a direct engagement of our actual selves in our imaginative processes in virtue of the similar cognitive mechanisms and brain structures involved in both actual and merely imagined experiences. Indeed, given the cognitive parallelism between imagined states and actual states of experience, there seems to be some kind of interdependence between the imagined and the actual self. On the other hand, such structural

---

[46] For reviews of such works see (Currie and Ravenscroft 2002; A. I. Goldman 2006; Kosslyn et al. 2006; Leeuwen 2016).

similarities are not sufficient to prove that the psychological profile of the self involved in imagination is not deprived of its descriptive content, as suggested by Nichols' theory and implied by Pessoa's insights on imagination. In fact, as the aforementioned example of the amnesiac patient waking up in the middle of the night with a headache made clear, the fact that the I-concept is descriptively empty – the amnesiac does not know who they are – does not preclude that they can still undergo experiences that can be self-ascribed – they still feel the pain of the headache and are able to self-ascribe the affective state they are in. Similarly, the engagement in imagination of neural and cognitive mechanisms that are similar to the ones involved in the real-life experiences does not prove the existence of a descriptively richer kind of self. Rather, it merely shows the occurrence in the imagination of representations marked by the I-concepts, without that implying that the imagined self is descriptively specified in any way. It may well be that the imagined self simply exploits the cognitive engine of the actual self. This concern is very similar to the one we previously raised in *Section 4.1* of the present chapter regarding the difference between the *processes* or *mechanisms* deployed for imagining being someone else and the *contents* of the imagining, i.e. what is actually imagined. We employ general cognitive processes that are assumed to be shared among humans. Said processes do not bear any substantial mark of our actual self, but merely the mark of the standard functioning of a humanly constructed cognitive architecture.

More decisive evidence comes from work done in social psychology and neuroscience. In a fascinating series of studies, Nicholas Epley and colleagues (Epley et al. 2004) showed how our perspective-taking tasks are accomplished by a process of "anchoring and adjustments" (Tversky and Kahneman 1974). This cognitive heuristic mediates our access to other people's perspective through our own first-person perspective. When mindreading others, perceivers tend to show the remarkable tendency to first *anchor* their evaluation on their own perspective, probably because it is usually the one that is most salient or evident to them. On the basis of this judgmental anchor constituted by their own first-person perspective, they then go on to *adjust* their initial evaluation based on an assessment of the self-other discrepancy at play. A higher degree of difference between the perceiver and the target would require a greater number of adjustments for the perspective-taking to be sufficiently accurate. Interestingly, Epley et al.'s results found neuroscientific validation in a later study run by Diana Tamir and Jason Mitchell (D. I. Tamir and Mitchell 2010) in which they found greater activity in the medial prefrontal cortex when perceivers were faced with cases of a wider self-other discrepancy. Moreover, both studies showed that people were significantly slower when making inferences about perspectives that more significantly differed from their own.

I take this research to show a psychological pattern that is consistent with the argument I have been supporting in this paper. When imagining being somebody else, that is, not only people tend to do it through the lens of their own first-person perspective, but they also need to "come to terms" with the perspective they are currently occupying – a point which is proven by the slower reaction times when assessing perspectives that more radically differed from their own. The anchoring effect seems to show that our first-person perspective constitutes the point of reference for our perspective-taking tasks.

A caveat is needed at this point. The studies I just mentioned do not deploy variables that perfectly map the conceptual granularity of the notions I am deploying in this chapter. The variables deployed by Epley et al. (2004) and Tamir et al. (2010) include notions such as "mentalizing", "mental states inference", and "perspective-taking", which are often used interchangeably and could therefore capture a variety of mentalizing strategies. This means that what these studies mean by perspective-taking does not perfectly map onto our notion of empathetic imagination.

There are some reasons to scale down these legitimate worries.

First of all, we are looking for empirical *support* for the claims defended in this chapter, not for a bulletproof empirical validation of them.

Second of all, these findings fit more intuitively with mentalizing strategies that are closer to processes of imaginative perspective-taking, rather than other ways of inferring other's mental states. For example, if people in the studies were basing their evaluations on a set of generalized assumptions about other people's psychology or of stereotypical characterizations, then it would be hard to explain their egocentric tendencies. Indeed, one of the desirable features of various kinds of psychological generalizations is that they hold constant across various contexts without varying as a function of (e.g.) self-other discrepancy.

Third, it seems that the tendency to anchor our imaginative representations to our own first-person perspective would seem all the more appropriate when we are interested in modelling perspectives through empathetic imagination, i.e. when trying to model first-person takes on certain events experienced by another individual. Since our knowledge about how experiences feel like from a first-person perspective finds its most exclusive source in our first-personal acquaintance with experiential tokens that instantiate the relevant phenomenological properties (D. K. Lewis 1990; Paul 2014), it seems natural to expect that their imaginative counterparts will substantially hang on (and, therefore, be anchored to) our own acquaintance with them. For instance, my knowledge concerning what it is like to have a child will dramatically depend on my first-personal acquaintance with the experience

of being a parent. Moreover, an eventual lack of that personal acquaintance would be likely to determine my inability to accurately conjure up an empathetic imagining of what it might feel like to be parents (Paul 2014).

Having (at least partly) downsized the worries about the significance of the studies mentioned earlier, we can now turn to review some further evidence that seems to speak for our case for a descriptively richer kind of self in imagination.

Particularly significant for this purpose is the literature concerning the interaction between emotion and imagination. Here we can see indeed that the way in which our cognitive and affective engines interact with imagination crucially depends on the kind of individuals we are, i.e. it bears traces of the kind of self we are.[47]

It is easy to ascertain in our daily life a crucial link between our perceptual states and our emotional states. Many instances of emotions, indeed, depend on a perceptual appraisal of certain features of our environment: I jump out of fear as a response to my seeing a spider walking on the pages of my book and I feel joy when looking at my lover's face smiling at me. In these simple examples, the perceptual states I happen to be in influence and ground my emotional responses. Such a direct link happens to remain intact even between imaginings and emotions. I am moved to sadness by Werther's unhappy story and I shiver in fear if I imagine a possible upcoming catastrophe.
This link between imaginative states and emotional states fulfils a strategically advantageous evolutionary function: it helps us navigate our environment more efficiently by enabling appraisals that are not limited to the here and now of our immediate perceptual space but opened up to the space of possibility. If I imagine that there might be some threats hiding in a dark street in a dodgy area of the city, then my consequent fear response is likely to wisely orient my behavior in order to avoid a potential threat.

Empirical work on this topic suggested that our affective responses seem to be an indispensable component of our decision-making processes and practical reasoning (Damasio 1994; Gendler and Kovakovich 2006; Leeuwen 2016). Indeed, a crucial ingredient of our practical reasoning depends on our capability to envision the possible consequences of our actions, in ways that allow us to assess the value of possible outcomes linked different behavioral possibilities. Given the action-guiding function of our emotions and the action-guiding function of our imaginative representations, an

---

[47] From now on, whenever I talk of imagination without any further specifications, I will mean a kind of imagination that is ascribable to what I called "empathetic imagination" or, simply, empathy.

interaction between the two faculties of human cognitive architecture is not only desirable but actually instantiated by neurotypical individuals. As perceptual states (e.g. seeing a spider) are likely to trigger emotional responses (e.g. fear) which subsequently influence the behavioral outcome of an agent, so do our imaginings: imaginatively simulated outcomes stimulate genuine emotional responses, or, as it is sometimes said, are "affectively encoded" or "somatically marked" (Damasio 1994). A fundamental feature of this link between affective states and imaginings is constituted by its *automaticity*. Our emotional responses seem to automatically raise from our imaginings. Sometimes it is impossible for us to silence a certain emotional reaction. No matter how hard I try, I cannot impede being disgusted by spiders. That is likely to constitute a major obstacle in my attempts to empathetically imagine the perspective of somebody who finds tarantulas cute, adorable creatures.

This has some crucial implications for the involvement of the self in our imaginative process. As in the case of spiders, my fear of these dreadful creatures tells something about *me*, namely it reveals something about my affective attunement to certain features of the world. This affective attunement is, as we saw, "imported" by my imaginings as well. In fact, what we happen to experience in response to a certain imagined scenario is likely to depend crucially on the kind of self *we are*.

Another persuasive example is that of sexual arousal: my ability to be aroused by a certain imagined scenario will crucially depend on my preferences such as my sexual orientation, my sense of erotism, etc. It is unlikely that I can come to first-personally appraise the sexually appealing features of an experience underwent by a target whose perspective I am trying to empathetically imagine if I do not share with him some fundamental preferences about erotism.

This implies, though, that my emotional responses and my overall affective attunement to the world, besides enabling my capability for perspective-shifting through empathetic imagination, is what is also going to lead it astray, preventing me to surf among an indefinitely large number of perspectives.

Before concluding, one final brief clarification is needed. The boundness of the imagined self to our actual self could generate the worry that our attempts to empathetically imagine somebody else's perspectives can be successful only as far as the other person is relevantly similar to us. That would deprive perspective taking of much of its capability of conveying a genuine insight into other people's perspectives. Empathetic perspective taking would amount to a mere exercise of self-projection in various contexts.

This is neither desirable nor entailed by the arguments presented so far. Indeed, the constraint constituted by our actual self can and in fact do get overcome. We can indeed, on some occasions, manage to imaginatively endorse a way of looking at the world that we would never recognize as

ours. Claiming that the imagined self parasitically depends on the actual self (as we did throughout the entire paper), does not rule out in principle the possibility of exploring alien perspectives. If the arguments presented so far though are on the right tracks, the issue concerning the possibility imaginative "migrations" towards relevantly dissimilar perspectives should be approached by acknowledging the necessity for our imagined selves to "come to terms" with the actual selves we are.

This set of issues, will constitute the main material for the next chapter in the present work in which I will try to focus on the surprising epistemic opportunities connected to the issue of the bounded self in empathy.

Onwards

Coming back to where we started, what Pessoa got wrong about imagination, specifically of imagining being someone else is that, to accomplish this task, we have to be "nothing", a mere mental vehicle whose psychological profile is so underdetermined that it can effortlessly surf among an unconstrained number of different perspectives.

As we saw, though, an empty self could not tell us anything about being someone else. As I tried to show, an imaginatively "thicker" self, one that bears traces of the actual self that produces it, is not only the irreplaceable requirement for empathetic imagination to take off and succeed, but also what exposes empathetic imagination to the constant risk of failures (e.g. imaginative resistance).

In the following chapter of the present work, I am going to address more in detail two specific ways in which the constraints posed by our actual self on our imagined self and therefore on our empathetic projects. I am going to focus on two very influential cases that have been largely influential in the specialized literature, namely Peter Goldie's critique of empathy as both unrealistically ambitious and irredeemably egocentric and L.A. Paul's account of the so-called transformative experiences.

Chapter IV

What Empathy Cannot Do: Transformative Experiences

1. Introduction

While out on a date with a friend, I might decide to buy ice-cream. I have reasonable expectations about how my ice-cream is going to taste: I can easily project myself in that future scenario and try to figure out what kind of flavor I will choose – e.g. vanilla ice-cream.

This prospective exercise is easily accomplished at least because two conditions obtain: (i) I know how ice-creams taste like, (ii) I know what my preferences are when it comes to ice-cream flavors. In other words, since I can imagine what it feels like to taste ice-cream, I can build expectations about how that experience will unfold and make informed decisions on the basis of my imaginative representation – e.g. I can pick a flavor.

Imagine now that I never had a chance to eat ice-cream. How would it be possible for me to have reasonably detailed expectations about it? I do not know what ice-creams taste like (i) and, therefore, cannot know what my ice-cream preferences might be (ii). As a result of this, I cannot adequately project myself in my future ice-cream-eating experience and make informed decisions on the basis of what I imagine. If I were to finally try ice-cream, I would acquire the relevant knowledge to build imaginative representations about ice-cream-involving scenarios and make informed decisions on the basis of my newly developed ice-cream preferences and my newly acquired imaginative skill. My experience with tasting ice-cream endowed me with new knowledge about certain experiential facts, i.e. the taste of ice-cream, and, as a result, it provided me with a specific value-function associated with ice-creams and ice-cream flavors, i.e. it has provided me with new preferences I did not have before undergoing the experience.

Speaking somewhat emphatically, we can conclude that in the latter case, the experience changed, in some very minimal sense, who I am. After my experience with ice-cream, I not only have acquired new knowledge, but I also added a new card to my pre-existing deck of preferences. From that moment on, my choices involving food and dates with friends will likely be influenced by my acquaintances with the taste of ice-cream.

Of course, the example of ice-cream is an overtly negligible and unremarkable case: the stakes are low and the personal changes determined by the experience in question are not important enough to

count as actual changes or *transformations* of the self. If we could subtract the experience of tasting ice-cream from the stock of experiences people had in their life, I am confident that most people would probably agree on the fact that it would not change them much. In other words, ice-cream-self is not so different from non-ice-cream-self: all of our core values would likely be left intact, and so would most of our beliefs, life goals, and ambitions.

At the same, though, it is easy to illustrate how the same logic described in the above example could be generalized to encompass higher-stakes cases in which the experiences we undergo have a much more profound impact on us.

In the contemporary philosophical debate, mostly thanks to the work of the philosopher L. A. Paul, these experiences have come to be known under the name of *transformative experiences* (henceforth, TE for the singular form, TEs for the plural). TEs are those kinds of experiences that follow the same logic described in the previous example about discovering the taste of ice-cream: they imply an epistemic transformation, in the sense that they provide us with knowledge about facts we ignored before undergoing the experience (e.g. the taste of ice-cream) and change us as persons, by e.g. providing us with preference we did not have before (e.g. preferences related to ice-cream flavors).

TEs have animated a flourishing philosophical debate as they succeed in raising some prominent issues about rational decision-making: how are we to make decisions involving transformative experiences if we do know neither what they feel like (i.e. their epistemic import), nor how they will impact us (i.e. their personal implications)?

As I will try to show in the present chapter, TEs represent a powerful case to highlight not only crucial issues related to decision-making but also some of empathy's most significant limits. The implication TEs on our capacity to perspective-take has been rarely tackled as such since the debate has tended to focus on the self-disrupting dimension of TEs and their implication for prospection and decision-making.

In what follows, I will mainly focus on issues pertaining to perspective-taking, rather than decision-making by asking questions such as: how can we imagine experiences we did not have? And, how can we imagine being kinds of people we are not, i.e. having preferences, values, dispositions we do not actually have? These are *intra-personal* issues as much as they are *inter-personal*: they have implications on both how we think of ourselves in the future and at the same time impact how we *understand* alien perspectives and individuals who are different from us.

I believe that the general concern raised by Paul could be restated in these terms: how is it possible to imagine occupying a perspective that is radically different from our own? That is, how is it possible to imagine oneself as a significantly different self? As we can see, it does not really matter whether the different self in question is simply a transformed version of ourselves at a different time, or a different person altogether.

TEs are a particularly interesting case for the purpose of the present work for they seem to deepen the point I raised in the previous chapter about the bounded self. The bundle of issues related to TEs could, indeed, be interpreted as highlighting the limits of our imaginative (or, more broadly, representational) capabilities due to the kind of selves we are. The stock of experiences we had constrains our capability to conjure up the relevant imaginings about kinds of experiences we did not personally undergo. At the same time, the kind of individuals we are – i.e. our preferences, dispositions, values, etc. – does not allow us to fully appreciate inhabiting a different self's perspective – with other dispositions, another system of value and preferences. In other words: who we are shapes and constrains our empathic processes. In what follows, I will explain more in detail how.

2. Epistemically and Personally Transformative Experiences: intrapersonal and interpersonal

Before delving into the challenge posed to empathy by TEs, it might be useful to say a couple more words about them so that we approach with a clear mind the constellation of issues related to them.

According to Paul, paradigmatic TEs are those that involve both an epistemic and a personal transformation. As anticipated in the previous section, epistemically TEs (ETEs) are those that provide you with new knowledge about some truths you could not have access to before actually having had these experiences. Personally transformative experiences (PTEs) are those that impact you *as individuals*, thus changing you "in some deep and personally fundamental way, for example, by changing your core personal preferences or by changing the way you understand your desires and the kind of person you take yourself to be" (Paul 2015: 476).

One of Paul's favoured examples of an epistemically transformative experience is Frank Jackson's super-neuroscientist Mary (Jackson 1982). As widely known, Mary spent her whole life in a black and white room, so she never had the experience of seeing colours. In spite of that, Mary is an amazing neuroscientist who knows everything there is to know about colour vision from a scientific point of view. Still, when Mary leaves her black and white room and sees red – or any other colour she happens

to look at – for the first time, she is acquainted with new knowledge about colour vision, she now knows *what it feels like* to see colours. When she leaves the room, Mary has an epistemically transformative experience in the sense that, thanks to her seeing colours for the first time, she now knows something new about the *experience* of seeing colours. The relevant distinction is the one between the so-called *descriptive* mode of presentation and the *subjective* mode of presentation (Paul 2017). Before seeing red for the first time, Mary knew everything there is to know about seeing red only in terms of a scientific description of it. Still, she was missing some facts about seeing colours that were accessible for her only from a first-personal point of view. The only way for Mary to achieve this specific feature of colour vision was experiencing it herself since "descriptions, stories, and testimonies could not teach her what it was like to see colour for the first time" (Paul 2014: 9).

In the original formulation of Mary's thought experiment, the focus was merely on the epistemic upgrade achieved through experience. Nothing was said about the possible life-changing consequences that this experience might entail in her case. But nothing prevents us from imagining that, after personally witnessing the mesmerizing grace of colours, Mary decides to quit her admirable career as a neuroscientist and become a painter. This would surely be not only an epistemically transformative, but also a *personally* transformative experience. Moreover, the change appears even more radical if we notice how Mary could not foresee it before actually undergoing the experience of seeing colours. Until that moment she could not even conceive of her undertaking a career as a painter.

Before moving on, it is necessary to clarify an aspect that will turn out useful in our subsequent analysis of empathy and TEs. When discussing PTEs, Paul tends to focus on the changes that occur to one's self-representation, which seems to be conceived of as a metacognitive awareness that something about oneself has changed. From that, one might draw the conclusion that paradigmatic cases of PTEs come with the ability of the subject to keep track of the radical changes that occurred to her. However, certain personal changes might occur without us explicitly realizing them. In other words, personal transformations can go unnoticed by the subject undergoing them. For instance, they might affect our character traits or dispositions, which are usually elements that do not figure in in the foreground of one's conscious awareness. Indeed, it is perfectly conceivable a case where a friend tells us something along the lines "you became so cynical" and where we experience her words as an unexpected epiphany about ourselves, something we suddenly came to realize only after our friend told us so. Now, we can imagine that such a change – viz. becoming cynical – occurred as a consequence of an experience or a set of experiences we underwent. Maybe the character-shift started to took place because we accepted a position as a finance broker at Wall Street: our constant exposure to cynical behaviours and the careless predicaments of our colleagues and bosses ended up

influencing our worldview. Still, this event did not significantly affect the self-image we keep of ourselves: we still believe that we are the same naïve person as we used to be before joining the Wall Street jungle. Taking stock, we can conclude that personal transformations can impact us in ways that elude the spotlight of self-awareness. This point should be kept in mind because it will turn out to be crucially relevant when we will come back to the challenge of imagining perspectives that involve different outlooks or personality traits. As we will see, the usual unconscious dimension of these psychological variables is what will significantly complicate the work of empathy.

But let us not come back to the case of TEs. As already stated in the previous section, the bundle of issues highlighted by TEs are not exclusively a problem of how we understand *ourselves* but also about how we understand *others*. Indeed, as TEs give rise to empathy failures araising from the impossibility of imagining radically different versions of oneself, then of course the same problem arises when confronted with different people altogether. In virtue of the epistemic and personal gap that separates us from different others we cannot imagine what it is like to be them.

Coming back to the case of Mary, if she does not know what it feels like for her to see colours, it is also true that she could not know what it feels like *for anyone else* to see colours. Before leaving her black and white room, she probably was not in the condition to fully understand the reasons that led individuals like Rothko or Yves Klein to produce paintings exclusively made out of colours. She could not even empathize with the aesthetic enjoyment that most people feel in front of those canvases.

However, what makes the issue of TEs so interesting is that they do not emerge only within the context of implausible thought-experimental scenarios. According to Paul indeed, our lives are recurrently shaped by these kinds of experiences.

Let us then consider a more concrete example. A real-life case of transformative experience explicitly addressed by Paul (2014, 2015) is that of becoming parents. She stresses how parents seem to be acquainted with experiences that are different from any other they had before (epistemic transformation) and, at the same time, seem to go through radical changes as individuals (personal transformation). In light of this, our intuitions convey this kind of judgment: it seems impossible for people who are not parents to empathize with the (transformative) experience of being parents. With Paul's words:

> If the salient details of the transformative experience of producing and becoming cognitively and emotionally attached to your child are epistemically inaccessible to you before you undergo this type of experience, then you cannot, from your first-personal perspective, imaginatively represent the relevant first-personal nature of the

> preference changes you will undergo. Because of your lack of experience, you lack the representational capacities needed to imagine, model, and grasp the nature of your future lived experience, and thus, of your future self. […] You cannot first-personally foresee or represent the new self you are making yourself into (Paul 2015: 474)

I take the conceptual bedrock of TEs as highlighting failures of our imagination – more specifically of our empathic imagination – when it comes to representing perspectives that radically differ from that which we actually inhabit in virtue of some major epistemic or personal gap that separates us from them. This allows Paul to illuminate some crucial issues related to decision-making: how is it possible to make informed decisions that are likely to involve a TE if we are not capable of accurately imagining what it is like to undergo them? How can we make decisions if we are not capable of first-personally assessing the relevant experiential features that constitute them? Said otherwise: how can I decide to undergo an experience if I have no clue about how it is likely to affect me?

I believe that these are fascinating questions well worth being asked and philosophically scrutinized. However, for the research purposes of the present work, I will largely leave them unexplored. I will rather focus on the consequences TEs have at the interpersonal level: namely, I will be interested in the implications that radically different perspectives have for empathy.

3. Epistemic transformation

It might be easier to start with ETEs as the conceptual pieces of the puzzle are already there and we simply have to put them together.

Throughout the present work, I stressed that empathy's main epistemic achievement be identified with *phenomenal insight*. ETEs seem, by definition, to preclude the possibility of phenomenal insight. This is due to the epistemic transformation that they bear to the subject undergoing them. Remember the case of Mary, her experience of seeing colours for the first time bears with it an epistemic transformation that she could not imaginatively envision before actually having that experience.

This phenomenon is usually discussed in the specialized philosophical literature as the 'knowledge argument'. Traditionally, the knowledge argument was advocated by philosophers in order to show how there might be a kind of knowledge about conscious experiences that cannot be reduced to or spelled out in terms of mere physical properties. Mary's knowledge of what it feels like to see colours cannot be achieved through her perfect knowledge of the physical processes underlying the experience of seeing colours – e.g. the brain's activity when experiencing colour vision.

In order to appreciate the persuasiveness of this argument in the context of empathy, we do not have to commit to any heavy-weight theory about the nature of consciousness. We merely need to be willing to accept the unsurprisingly uncontroversial claim that in order to know how certain experiences feel like, we need to rely on our personal acquaintance with experiences of the same *kind*. What does that mean precisely? As I tried to show with the example of the ice-cream flavours in the previous section, in order to know what it is like to taste vanilla ice-cream, I need to have tasted vanilla ice-cream at least once in my life, and, of course being able to *recall* what it felt like to taste it. In the absence of that relevant experience, I cannot possibly imagine the experience of tasting vanilla ice-cream.

However, one might argue, what if I know the taste of vanilla, even if I experienced it in a different form from ice-cream like, say, a vanilla cream? One might say that there is a subtle sense in which *every* experience is different from anyone else we happened to undergo in the past. Even if I drink coffee every morning right after waking up, there is a sense in which the coffee I drink does not *exactly* taste as it did the morning before. However, it would be uselessly pedant and epistemically unfruitful to claim that I do not know what it feels like to taste coffee only because I cannot be acquainted with the *whole* spectrum of the possible coffee-tasting experiences. The subtle differences between my repetitive experiences of drinking coffee in the morning are way less relevant than the commonalities they share in the process of knowing what it feels like to taste coffee. Here is how Paul discusses this issue:

> The problem posed by epistemically transformative experiences for decision-making is not a problem that arises in virtue of the fact that *any* experience may—in some way—be qualitatively new or different from your past experiences. For example, you may have tasted ripe pineapple in the past, and so you know what the experience of tasting ripe pineapple is like, broadly speaking. But you may not know what it will be like to taste *this* piece of pineapple in front of you. Perhaps it will be ever so slightly acidic. Or perhaps it will be cloyingly sweet. Perhaps as a result it'll taste a tiny bit different from any other piece of ripe pineapple that you've had in the past. Does this fact create a problem for your breakfast decision? No. Epistemically transformative experiences arise from having new *kinds* of experiences, not from new token experiences that are instances of the kinds of experiences you already know about. […] Minor changes or variations in properties that are not the dominant, kind-defining properties of the experience are not relevant to knowing what that kind of experience is like. (Paul 2014: 36)

Those "minor changes" that intrinsically characteristic of any reiterated experience are not "kind-defining properties", i.e. their epistemic magnitude is not significant enough to determine a *change* in the kind of experience a subject is undergoing. The emphasis on ETEs being about new *kinds* of experiences is what safeguards Paul's arguments from the picky triviality of the argument about the incommensurability of every token-experience.

However, one might ask another question along the lines of the previous one: when can a subject legitimately claim that their knowledge about a certain kind of experience is sufficient to achieve phenomenal insight about a certain instance of it? Indeed, it is clear that kinds are not fixed, and one can move upwards and downwards across different degrees of generality.

To better assess this concern, it might be useful to imagine, to the extent to which this is possible, a case of someone who never experienced the emotion of joy in their life. It seems like we can safely claim that this individual, call him Joeyless, is not acquainted with a certain *kind* of experience, namely the experience of joy. In light of Mary's example and Paul's argument, we can conclude that Joeyless cannot imagine what it is like to experience joy and, therefore, cannot empathize with anyone's experience of joy. However, one might think that "joy" is just one member of a general kind that we can call "emotional experience". In certain contexts, we might want to carve experiential states in a more coarse-grained fashion, like when we individuate the general kind of "emotional experience" to clearly distinguish it from other general kinds of mental states, such as "perceptual experience". So, now the question is: how do we determine the level of specificity and detailedness required to achieve empathy and phenomenal insight?

Unfortunately, there is no straightforward answer to that question. Remember when we discussed Cappelen & Dever's account of empathy: there is no *golden rule* that can determine *a priori* what counts as a sufficiently accurate representation of another person's perspective. The standards of accuracy for empathy are highly context-dependent: they are set by both the epistemic ambitions of the empathizers and, eventually, what we might call the "empathetic demands" of the person with whom we are empathizing with. As I will try to show in the next chapter, discussions about empathy are usually framed in a way that seems to take it to be something that one does in complete isolation and exclusively in dialogue with themselves. Empathy is conceived along the lines of the experience of reading a book or of daydreaming, activities that people usually do in solitary, letting their imagination run while quietly sitting on an armchair. However, central instances of empathy happen in the much more dynamic context of human interactions. In more ecologically valid contexts, our imagination is shaped, guided, and constrained by numerous contextual cues such as testimony, perceptual cues, etc., and it usually requires the constant updating of relevant information we take into account while perspective-taking.

Once we come to appreciate this more embedded dimension of empathy, we also come to appreciate how the standards of accuracy can vary according to the different contexts in which the empathic processes occur. In some instances, we might not need a very detailed and fine-grained understanding

of the experiential states our target is undergoing, in some other cases, an in-depth and detailed appreciation of the other's perspective might be so important that empathizers can be legitimately blamed for being too superficial in their understanding. Adam Morton has the perfect example to illustrate this point:

> Consider for example a person, Melanie, in an unusual and delicate trap. She has encouraged George to become attached to her because Eric has jilted her. Now George has suggested marriage and she might be favourably inclined since she has come to appreciate his straightforward affection and his undeviousness. But Eric has just contacted her and intimated that he has made a terrible mistake and would like to get together again. Her time with George has made her realize quite how devious Eric is, and she is far from sure that Eric is not just trying to mess with George or her, and will be off again once the damage is done. But she retains a deep longing for him. Does she love him? Not really, though she finds him exciting. She meets her friend Helen, who sees how upset she is and asks about it. Helen's reaction on being given the barest outline is, "Oh you poor thing; these hard decisions can really take it out of you. Melanie snaps at her and changes the topic. (Morton 2017: 183)

Morton notices how Melanie can rightfully resent her friend Helen for being so superficially empathetic. In a way we can certainly say that Helen has a fair understanding of the *kind* of situation Melanie finds herself in, i.e. she is in a highly distressing situation, she has to make a difficult choice where the stakes are exceptionally high and one option is not clearly better or worse than the other. What needs to be stressed again at this point as crucially important is that these standards are set by the nature of the specific social dynamics in which they occur. It is virtually impossible to establish *a priori* what might count as a satisfying standard of accuracy for empathic understanding.

The above considerations allow us to keep in mind that Paul's argument concerning differences in *kind* between various experiences relies on both our intuitive understanding of what makes one experience different *in kind* from another and the contextual nature of empathy's accuracy standards. On the one hand, we can say that virtually anyone is capable of detecting kind-defining properties that help them differentiate between various experiential states. To put it roughly, we can tell apart the experience of joy from the experience of sadness, the experience of tasting coconut from the experience of tasting pineapple, etc. On the other hand, we also know that what counts as a kind-defining property will be subject to a considerable degree of contextual variation.

Limiting our attention to the case of empathy, how fine-grained our empathic representation of another person's perspective needs to be will be defined by the specific nature of the empathetic task we are up to. In discussing issues that are partly independent from the arguments I am making here, Paul herself offers a good anchor to better illustrate the point I am trying to make:

> The distinctiveness of "what it's like" consists in its dynamic, active, forceful feel. While different phenomenal characters exhibit different characters (there is a felt difference between a token experience with the phenomenal character of seeing red and a token experience with the phenomenal character of seeing purple), there is also a felt dynamic quality or "oomph" that is common to each experience, in virtue of it being an experience with phenomenal character *simpliciter*. This felt resemblance or commonality is what makes an experience with the phenomenal character of redness and an experience with the phenomenal character of purpleness, and all other experiences with phenomenal characters, members of the *experiences with phenomenal character* kind. (Paul 2017: 206)

We can certainly claim that experiences can be grouped into different kinds according to whatever sorting principle we decide to refer to. In the above passage, it is certainly true that we can individuate a specific *kind* of experiences by referring to a sorting principle that individuates them according to their *phenomenal character*, i.e. by the fact that there is something that it is like to undergo those experiences. However, even if there is a kind-defining property that a wide variety of experiences have in common, ranging from the experience of seeing red to the experience of giving birth to one's child, it is also straightforwardly clear that in most contexts in which we deploy empathy that kind-defining property is very likely to be too broad to be truly informative about another person's perspective.

However, we can still conceive of cases in which such a coarse-grained empathic understanding of someone's perspective might be crucially important. We can indeed imagine a case in which the mere ascription of experiences with a generic "phenomenal character" might be a satisfying empathetic contribution to the proper understanding of certain circumstances. Consider the case in which we are trying to decide whether euthanasia might be the best solution for someone who is in a severe vegetative state. Now, whatever considerations we make and whatever conclusions we come to, being able to understand whether or not the person in question is capable of undergoing experiences that feel like something, is certainly a crucial element to take into account when trying to carefully assess the situation in question.

We can take another example to show how inappropriate a general coarse-grained grasp of certain experiential states might be for some other empathetic purposes. Namely, we can immediately understand why Mary cannot say she can empathize with the experience of seeing colors because she is acquainted with the general kind of experience known as "visual perceptual experience". She cannot reasonably say: "I know what it feels like to have visual experiences, therefore, I can imagine what it is like to see color, as seeing colors and seeing objects belong to the same kind of experience, namely a visual experience". Her knowledge about other perceptual states – even within the same sense modality – that are not precluded to her would be too poor of a guide in bringing her to

appreciate the nature of what it is like to see colors. Similarly, Helen's general understanding of Melanie's hard time in making the right decision cannot possibly count as satisfactorily, as the specificity of the social dynamic in question requires a much more fine-grained understanding from Helen.

George Elliot wrote "until we know what has been or will be the peculiar combination of outward with inward facts, which constitute a man's critical actions, it will be better not to think ourselves wise about his character" (*Adam Bede*, ch. 29). What should be added is that "the inward and outward facts" that are to be known in order to make a person "*wise*" about – or, as I would put it, to *empathetic* towards – someone else are set by the context. In some cases merely knowing how painful it is to end an important relationship is enough to empathize with our friend who just broke up, whereas in some other cases a much richer body of information will be required in order for us to be up to the difficult task of empathy – e.g. knowledge about how long the relationship was, how it ended, why it ended, what the emotional dispositions of our friend, their tendencies, commitments, etc. might be crucially important in order to properly appreciate what it feels like to be in our friend's shoes.

With this in mind, we can take stock of what I have been saying so far in the chapter. Paul's main point consists in claiming that relevant knowledge about certain kinds of experiences can be achieved only via acquaintance. A lack of experience necessarily results in a lack of empathic insight. This lack of insight generates the decision-making *impasse* Paul's 2014 book brilliantly illustrates. Furthermore, as I tried to show in the previous sections, this has major implications for the way in which we understand other people as well. Specifically, imagination cannot bring you the epistemic transformation that only the experience itself can give you. Stated with the terminology with which we familiarized in the previous chapter: our imagined self inherits the experiential – and, therefore, representational – deficits of our actual self. This is what impedes the imagined self to explore perspectives that involve the epistemic transformation which it lacks in virtue of its dependency on – which I called its *boundness* to – the actual self. Our set of lived experiences is what pre-determines the perimeter of our empathic insight: our stock of personally lived experiences set up the limits of our interpersonal insight.

The inevitable consequence of all this is that, in virtue of such a fundamental gap in their knowledge, empathizers cannot achieve phenomenal insight into ETEs unless they personally underwent them at some point in their life.

4. Personal transformation and Goldie's anti-empathy

In this section, I will try to isolate two kinds of issues that stem from Paul's arguments about PTEs, one that has to do with the limits of humans' foresight, and another that is more similar to the problems we highlighted in the previous sections about ETEs, having to do with genuinely *epistemic* issues related to the representational limits of our imagination when it comes to achieve an insight into perspectives that are considerably different from our own.

As a starting point, it interesting to compound Paul's considerations about TEs to some insightful remarks made by Peter Goldie regarding the conceptual impossibility of empathetic perspective-taking (Goldie 2011).

As already anticipated in the previous chapter, Goldie endorses a clear-cut distinction between imagine-self, which he calls in-his-shoes imagining, and imagine-other, which he calls empathetic perspective-shifting. This distinction is the foundation of his critique against empathy. We do not need to clarify this distinction because it has been largely clarified in the previous chapters. Suffice it to remind us that in the first case we imagined-self and our actual-self coincide (to a large extent), whereas in the second case the imagined-self is taken to be centred on the *other* rather than on ourselves. In the first case, the imagining is essentially self-referential, whereas in the second case it is – or attempts to be – essentially other-directed.[48]

This distinction is crucial because it is what grounds Goldie's claim about empathy's conceptual impossibility. For him, when deployed as a strategy to understand targets that are too different from the empathizer, empathy is doomed to failure. One case discussed by Goldie is of particular interest for us, namely one in which the empathizer's and the target's character traits and psychological disposition differ.

[48] Walton (1990: 28) famously claimed that, in some sense, all imagining is self-referential. By this he meant that, when trying to imagine doing or experiencing things – as opposed to merely imagining that (or belief-like imagination) – we export much of our selves in the imaginative project. I will not engage with this view in the present chapter for two reasons: (i) it will bring us to far afield from the present concerns, (ii) in the previous chapter I already explored issues of a similar sort, when I tried to defend the view according to which, with empathy, what we imagine is shaped by the kind of self we are.

4.1. Dispositions as the background of first-personal agency

In his earlier analysis of empathy, among the various features needed for empathy to be successfully instantiated, Goldie mentioned what he called a "substantial characterization" of the other person (Goldie 2000). By this he meant that in order to "centrally imagine someone" – viz. empathize with them – we need to have at least some relevant knowledge that could help us portrait the psychological profile of our target, e.g. some knowledge about the other's values, desires, emotional dispositions, etc. These diverse aspects of someone's psychological outlook are usually gathered under the labels of "character trait", "psychological dispositions" or, simply, "personality".[49] These compound psychological portraits encapsulate a variety of features that can be ascribed to other people and easily deployed to frame our understanding of them. When we say that someone is a "kind" person we are indeed making several claims about them. At the very least, we are claiming that they tend to have certain kinds of desires, e.g. they are likely to have the desire to help someone in need, behave in a certain way, e.g. they are likely to help someone in need, and undergo certain kinds of emotional states, e.g. they are likely to feel compassion for someone in need.

Talks of personality, character traits, psychological dispositions could turn out to be a very useful shortcut for explaining, predicting, and understanding someone's mind and behavior. Goldie illustrates this point with a couple of enjoyable examples:

> [W]e use trait discourse to help us to predict things about people. Adrienne advertises in the lonely-hearts for a kind, confident, intelligent man because she wants a man with these traits, and because she thinks that a man with these traits will reliably think, feel and act in a kind, confident and intelligent way. […] So we describe, judge and predict by using personality discourse. And finally, we use personality discourse to explain. Freddie is your new boss. You go into his office to ask him a question, and he shouts at you and humiliates you. After crawling out, with his horrid words ringing in your ears, you ask someone by the coffee machine why Freddie should have done such a thing. 'Because he's a bully', comes the reply. You feel better. (Goldie 2004: 5)

---

[49] I will leave aside the definitional issues related to these quasi-technical terms. Indeed, one could rightly argue that character traits, psychological dispositions, and personalities are not the same thing. For instance, one's disposition to be afraid of spiders does not amount to a personality and could not even be properly defined as a character trait. These notions are admittedly blurry, but we do not need clear-cut, rigorous conceptualizations of them in order to properly assess the issues we are dealing with in this chapter. In the present context, they should just be taken as generalizations about other people's psychologies that compound several variables about their thinking, feeling, and behavioral tendencies.

In our mindreading routines, personality and character trait attributions abound. They offload some of the cognitive effort required to figure out other people's mental states and attitudes and optimize our understanding of other's minds and behaviors.

At this point, it might be useful to retrieve a useful distinction recently drawn by Shannon Spaulding between different "goals" and "approaches" of mindreading (Spaulding 2018). For instance, when our mindreading *goal* (among other possible ones) is "efficiency", that is a fast and cognitively "cheap" understanding of the target that allows us to minimally navigate a particular social dynamic, or the "confirmation of our own worldview", i.e. the framing of the situation within a perspective that looks to some extent "familiar" to us, we tend to *approach* mindreading in a relatively effortless way: we rely more on cognitive shortcuts such as social categorizations, i.e. we sort people into social categories. For instance, if I categorize someone as "a Wall Street financial broker" I am likely to frame my understanding of their behavior in a way that is consistent with the features that are usually ascribed to that specific social category (e.g. egocentric, profit-oriented, etc.). These help us "characterize" individuals that we take to fall within a certain category. Sure, social categorization, as basically every kind of generalization, can lead us astray (I am sure the world is full of kind, altruistic, and compassionate Wall Street brokers) but it is undeniable that they can work just fine in providing us with a fine understanding of at least some social dynamics. As rightly noted by Elisabeth Camp (2017: 7), characterizations (of the sort that are produced by social categorization), do not commit us to believe that their subjects should mandatorily possess those features, i.e. they do not have an actual normative force. They merely facilitate and structure our interpretations by presenting a certain bundle of features as particularly *fitting* for the members of a particular class.

If we focus on their mere epistemological upshots,[50] social categories – as a kind of characterization – help us frame and organize the information available about a certain person by giving us a pre-compiled "reading key" of the other.

Similar considerations can be made for personalities, character traits, and psychological dispositions. They do not mandate the ascription of certain mental states (e.g. kind persons can and indeed do behave unkindly sometimes), but they certainly do facilitate the interpretative work by structuring our understanding of the other in a way that is functional to efficient mindreading. Personality traits can work as a rough and ready explanation for a variety of social instances. They help us narrow

[50] And leave aside, for the moment, the possible e.g. moral upshots of sorting people into social categories which could be rightly be considered as being dangerously close to social discrimination.

down the relevant information about the other and frame it in a way that serves prediction and explanation.

Equipped with these conceptual distinctions we can now turn back to the problem of empathy and to Goldie's critique. According to him, when the goal of our mindreading is empathy, then trait-attribution is not an available approach. This is due to both the epistemological prescriptions demanded by a process such as empathy and the very nature of trait-attribution. As it should be clear by now, empathy is a complex, demanding exercise of mental mimicry that aims at giving us back insight into a perspective other than our own.

Now, let us see how these epistemological demands imposed by empathy couple with a trait-attribution strategy. When we try to understand someone through the lens of psychological traits, we basically categorize them as belonging to a certain class where certain psychological features are likely to be instantiated. For example, suppose that our friend Andrew is a kind person and we are confronting ourselves with his experience of stumbling on an old lady struggling to cross the street. Now, if we were to predict his behavior, we could resort to our knowledge of him being a kind person and confidently predict that he will likely help the old lady, as this is what we would *expect* from kind people. Similarly, if we are in search of an explanation as to *why* he actually helped the lady cross the street, we can again resort to trait-attribution, i.e. he did it *because* he is a kind person and that is how kind people usually behave in similar circumstances.

However, we might want something more than merely predicting or explaining the other person's behavior, that is, we might want to empathize with Andrew's experience. We want to imagine what it was like to occupy his perspective when he saw that old lady and decided he was going to help her.

At this stage, Goldie's critique of empathy is crucially important to understand what can hinder empathy. Assume that I have psychological traits that are relevantly similar to Andrew's. I am also a kind and caring person and with respect to this particular aspect, the gap between Andrew and me is inexistent or, better, so small to be unimportant. Therefore, if I were to empathize with Andrew's decision to help the old lady cross the street, I would just have to imaginatively place myself in his situation and "see" how I would feel: since I am relevantly similar to him, I would have relevantly similar feelings, thoughts, and tendencies and I would feel the urge to help that poor old lady. The absence of relevant *gaps* between Andrew and me – at least with respect to the situation in question – is what allows a smooth shift between my actual first-person perspective and his.

Now, imagine that instead of being a kind person, I tended to be quite careless and selfishly oriented in my interactions with others. In this case, if I were simply to place myself in Andrew's situation, I would likely produce an imaginative representation that relevantly differs from Andrew's actual perspective. I would not feel that caring, spontaneous instinct to help the lady but I could rather experience indifference. The personality gap between Andrew and me would obstacle my empathic understanding of his perspective.

What makes Goldie's argument so interesting is that, when it comes to empathy, this gap cannot be bridged by our mere knowledge of it, i.e. by means of a trait-attribution strategy. Indeed, if I were to explicitly factor in Andrew's kindness when trying to imagine his perspective, then I would bring on the *foreground* of my awareness something that is only implicitly present in the *background* of Andrew's first-person perspective. In other words, my knowledge about Andrew's kindness is likely to interrupt the empathetic imaginative flow since I cannot fully identify with it. Knowing *that* Andrew is a kind person is not enough to be able to represent what it is like, for a kind person, to act out of kindness. Kindness is not a property that is usually explicitly represented *as such* by kind people behaving kindly.

Indeed, it is unlikely that Andrew decides to help out the old lady with the explicit intent to *conform* to the ideal of kindness. Namely, it seems preposterous to postulate that his behavior results from the prescribed compliance with a proposition such as "I am a kind person".[51] Kind people are usually *disposed* to act kindly but do not act out of an explicit prescription of kindness. Goldie illustrates this point very clearly:

> [W]here the full-blooded notion of agency comes into play, A's attempt at perspective-shifting to B's psychology will have to involve taking on those aspects of B's characterization which differ from her own whilst at the same time not being conscious of them as such. The reason for this is that the typical role of these dispositions is passive or in the background in the sense that our conscious thoughts and feelings that feature in our deliberations are shaped by, but are not directed towards, these dispositions. The kind person has kind thoughts and feelings, but these thoughts and feelings are not typically about the disposition of which they are an expression. The person who loves his spouse seldom thinks that he is doing loving things because he loves her; and for the modest person, as Bernard Williams says, 'it is a notorious truth that a modest person does not act under the title of modesty'. (Goldie 2011: 309)

[51] By this I do not mean that such cases are still possible. Maybe someone wishing to improve their "moral skill" might actually decide to conform to an explicitly represented and externally imposed rule of kindness.

This means that, when traits differ across individuals, we are required to explicitly factor in the interpersonal equation the relevant differences. This taints the imaginative process with the additive implementation of a trait-attribution strategy. As long as this knowledge about diverging characteristics between target and empathizer endures on the foreground of the empathizer's awareness, their attempt will fail: they will not be able to grasp what it is like to occupy a perspective where these dispositional features are instantiated.

4.2. Goldie's critique and (personally) transformative experiences

Goldie's critique can be particularly enlightening when coupled with Paul's philosophical exploration of TEs, more specifically of PTEs. Indeed, when it comes to personal transformations, we can partly re-frame the issue as a major change in character, personality, or psychological disposition. Now, if we are trying to empathize with radically different perspectives, viz. those that would require us to go through some major epistemic and personal transformations, it seems that we would fail to imagine them at least partly in virtue of the arguments put forward by Goldie. When it comes to PTEs, indeed, personal transformations impede us to imagine what it is like to inhabit a perspective in which those core self-defining characteristics have dramatically shifted.

By placing Goldie's analysis Paul's considerations on PTEs in dialogue I am certainly not claiming that the personal transformations as defined by Paul are completely captured by what we called with Goldie changes or differences dispositions and personality traits. Of course, there is more to personal transformations than *mere* changes in personality traits. At the same time, as we saw in the previous sections, Goldie's analysis provided us with useful insights for understanding *why* it could be so difficult (if not impossible) to imagine radically different perspectives. Major interpersonal differences seem to force us to hold what we might call an "intellectualistic" take on the perspectives of others which can considerably hinder empathic insights. This I take to be the core of Goldie's message the can be deployed to further illuminate issues arising at the intersection between PTEs and empathy.

Before moving on, though, there is an ambiguity about the very notion of PTEs in its relationship with empathy that needs to be solved. It has to do with the issue of determining what exactly determines the unimaginability of personal transformations: is it an intrinsic limit of our representational capabilities or is it a structural feature of personal transformations? Namely, is the representational barrier posed by PTEs similar to the one we experience with epistemic transformations or closer to the essential unpredictability of certain future changes?

We need to highlight two ways in which it is possible to conceive of PTEs that have not been usually brought up in contemporary debates. We can assess the problem of PTEs either as a – particularly problematic – failure of human foresight or we can conceive of them along the lines of ETEs, i.e. as a genuine failure of human imaginative representational capabilities that is orthogonal to issues of foresight. In the first case, PTEs mainly highlight the troubling consequences for rational decision-making regarding life outcomes that involve personal transformations. In a nutshell, due to the complex and largely unforeseeable changes that certain experiences bear with them, we cannot rationally assess whether to undergo them or not, simply because we are incapable of anticipating *what changes* they will bear with them. In the second case, the problem is not much about our capability of foreseeing how an experiential outcome would unfold, but rather about our capability to *first-personally represent* the changes that are likely to be involved. Let me be clearer about it.

In order to assess the impact of personal transformations on our representational capabilities we first need to individuate two different *temporal-directions* of empathy, viz. predictive or retrodictive. In the first case, we are interested in foreseeing a possible experiential outcome, in the second case we are interested in reconstructing an experience that already occurred.[52] Besides that, we have to further distinguish among two possible *target-directions* of our empathic focus, viz. the others (other-directed) or ourselves (self-directed).

The most commonly debated cases of empathy are those that involve other-directed retrodictive and predictive uses of empathy, but Paul is specifically interested in cases of what she calls "prospective empathy for future selves" (Paul 2017), that is predictive self-directed empathy. She focuses on cases where a person P at time t1 wants to know what it would be like for her to have a certain experience at time t2. As already said, it seems legitimate to extend Paul's considerations on prospective empathy for future selves to cases that involve agents other than ourselves and retrodictive uses of empathy. Remember the case of Mary: not only could she not know what it was like to see red for her, but also for anybody else.

---

[52] One might argue that we often empathize with experiences that are *currently* happening in the present moment. That would be a fair observation, but I believe that those cases can be substantially equated with cases of retrodictive empathy. Indeed, in such cases, we are not trying to *foresee* what the target is likely to experience in the future based on the current evidence we have about the situation, but trying to *represent* a certain kind of experience, based on the information that is available to us. This latter case is structurally more similar to retrodictive empathy than to predictive empathy, as it centrally involves imaginatively re-constructing an experience rather than forecasting it.

In her 2014 book, Paul's introductive example of a transformative experience was the imaginary case of becoming a vampire. With this example, Paul states that it would be impossible for us to rationally decide to become a vampire because we would not know what it would be like to be a vampire.

> "Imagine that you have the chance to become a vampire. With one swift, painless bite, you'll be permanently transformed into an elegant and fabulous creature of the night. As a member of the undead, your life will be completely different. You'll experience a range of intense, revelatory new sense experiences, you'll gain immortal strength, speed and power, and you'll look fantastic in everything you wear. You'll also need to drink blood and avoid sunlight. Suppose that all of your friends, people whose interests, views and lives were similar to yours, have already decided to become vampires […]. They say things like: "[…] It's amazing. But I can't really explain it to you, a mere human—you have to be a vampire to know what it's like." (Paul 2014: 1)

Her impossibility for predictive self-directed empathy derives from the impossibility for accurate other-directed retrodictive empathy. Now, we have to ask whether, when it comes to personal transformations, the unimaginability is merely due to the intrinsic unpredictability of certain future outcomes or due to the very nature of this change.

For instance, when I try to imagine myself as a parent and I try to contemplate what kinds of personal transformations I might undergo – leaving aside for the moment the straightforwardly epistemic issues concerning the possible range of qualitatively novel experiential states –, is it just that I do not know *what kinds* of personal changes will eventually occur or that I cannot subjectively represent *undergoing* them?

The first problem concerns the mere uncertainty regarding the impact that certain intense experiences such as parenting can have on our personality. I simply cannot know for sure what would happen to me if I were to become a parent. Will I become an affective father? Will I start enjoying nights out with friends less? Will I develop an unbalanced ambition for my child's achievements? These are all changes that cannot be accurately forecast, at least not by most average human forecasters dealing with these kinds of life changes.

But then if it is other-directed and retrodictive, can I imagine first-personally embedding such changes? In other words, can I empathically imagine having those new dispositions and personality traits? If I cannot, then the problem is not merely linked to the unpredictability of my future, but actually to the failures of our imaginative skills. In a nutshell, it is largely linked to the set of issues brought up by Goldie's critique of empathy: we simply cannot accurately imagine having such traits, if they are not ours.

Before moving forward, it should be noted that the most obvious and appropriate answer to this question is probably one that acknowledges the co-occurrence of both issues: the limits of both foresight and empathic imagination. However, this answer would be rather cheap and would not allow us to adequately explore the philosophical implications of TEs. Indeed, if we were to reduce TEs to one case amongst others in the vast ocean of human foresight's fallibility, then the arguments about empathy's limits that rely on TEs would lose much of their strength. They will boil down to the mere acknowledgement that, *de facto*, humans can be very bad at forecasting certain experiential outcomes.[53] Sure, this does play a role that should not be neglected especially when dealing with issues linked to decision-making. However, these arguments do not constitute *per se* a particularly relevant challenge to empathy, as they could be easily dismissed as a mere problem of information-availability: if we were to know *what kind* of changes would occur, then we would have no problem in imagining what it would be like to undergo them. Goldie's critique of empathy is precisely what allows us to frame PTEs in a way that is relevant for empathy's representational possibilities rather than as a mere problem of prospection.

For instance, when you are trying to imagine yourself as a parent, you are asked to imagine a change that might involve e.g. your child as the most important priority in your life. Now, supposing that I am disposed to be my first priority and tend to have a fairly egocentric orientation in the way in which I plan and live my life, this major change in psychological disposition will likely be hard to imagine. If Goldie's analysis is on the right track, I would fail to imagine what it feels like to live by an all-encompassing feeling of care for my child. I can certainly know how people embodying that disposition do actually behave but I cannot appreciate what it feels like to *inhabit* such a perspective.

On the basis of this, we can certify that Paul's challenge against empathy does not merely work in the case of prospective empathy but encompasses the entire representational spectrum of empathy. Indeed, in light of Goldie's critique, we can reframe the challenge of PTEs in a way that cannot be accommodated as a coincidental feature of information-availability. If indeed an infallible predictive machine were indeed to produce a perfect prediction of the kinds of personal change we would undergo once decided to become parents, we would still fail to first-personally appreciate those changes and we would be left with the same representational impossibility that is at the basis of the

---

[53] Up to now, the best and most entertaining introduction to the issues of foresight is to be found in Daniel Gilbert's 2006 book (Daniel Gilbert 2006). This book brilliantly illustrates some of the most pervasive and predictable biases riddling human imagination and, therefore, humans' capacity to accurately forecast experiences they might undergo.

decision-making dilemma posed by TEs. We would fail to imaginatively identify with those traits and appreciate how they would play out in the concrete unfolding of our lives.

4.3. Resisting Goldie's critique: Imagining under constraints and the use of propositional knowledge for empathy

One might have good reasons to be baffled by the implausibly high epistemic demands placed on empathy by Goldie's account of empathy.

Some authors (e.g. Tietjens-Meyer 2015) have in fact stressed that, if we want empathy to achieve a seemingly flawless imaginative identification, then by its own definition, empathy turns out to be conceptually impossible. With Tietjens-Meyer's words:

> In excluding propositional thinking from empathy, he [i.e. Goldie] turns empathizing into something like a waking dream in which you're possessed by someone else's identity and a story unfolds from within that perspective. I appreciate Goldie's insistence that empathizing is a demanding imaginative activity. But, outside debates in philosophy of mind, I fail to see what is gained by treating the process of developing a characterization of another person as a precondition for, but not a constituent of, empathy. (Tietjens-Meyer 2015: 154)

Coming back to the case of personality traits and psychological dispositions, our knowledge about how these psychological features unfold in the context of human agency ideally takes the form of a generalizable set of propositions that inform us about some behavioral and mental features that are likely to be instantiated, in certain given circumstances, by people having them. In other words, we seem to have some sort of tacit "theory" or generalization about people's minds and behaviors. Again, my knowledge of how kind people usually think and behave leads me to expect certain kinds of behavioral, thinking and feeling patterns from people whom I take to instantiate the attribute of kindness.

Now, one might think that this knowledge can fruitfully work in tandem with empathy, even if we fail to perfectly identify with the psychological features that we are ascribing to our targets. These concerns should remind us of some issues I mentioned in the previous *Appendix*, when I briefly discussed the interaction between different kinds of imagination. In that section, even if I distinguished among different kinds of imaginings to pinpoint some crucial epistemic issues linked to different kinds of cognitive (i.e. imaginative) processes, I readily admitted the possibility of a tight interplay between them in order to bring about the relevant representations these imaginings are after. Similarly, we can say that the empathic imagination can rely on propositional knowledge – or, if we prefer, on a dose "theory" – to achieve insight into other people's perspective. In our specific case, it

can rely on sets of generalizable psychological characterizations in order to get a grasp on other people's minds.

These worries have been recently tackled by some relevant work in the philosophy of imagination. Specifically, Amy Kind has recently investigated the way in which imagination might lead to knowledge (A. Kind 2016).

Kind sets her theoretical move against the philosophical truism according to which imagination cannot, by definition, lead to knowledge. She reminds us that this widespread philosophical assumption is based on at least three features that are usually ascribed to the imagination. Outputs of imagination are indeed usually taken to be i) voluntary, ii) unconstrained by the actual world, iii) already known from the start by the imaginer. Let us take a very simple instance of imagination: my imagining of a winged horse. This imagining is:

i) *Voluntary*, in the sense that I *decided* to bring it about. I wanted to conjure up in my mind an image of a winged horse and that sufficed to generate the relevant imagining. One might contrast imaginings with percepts. Differently from imaginings, perceptual states are usually not under our voluntary control: my intention or desire to see the *Tour Eiffel* in front of me does not *per se* suffice to bring about the required percepts in front of my eyes.

ii) *Unconstrained* by the actual world because it is utterly unaffected by the fact that there are no actual winged horses in the world in which we happen to live. Imaginings do not represent the world as it actually is. In this case as well, a contrast with percepts makes this feature more salient. Percepts are *constrained* by the world in the sense that they usually depend – at least partly – on how the world actually is: it is because horses *do* actually have four legs that I perceive four legs when I encounter one. Moreover, differently from imaginings, percepts represent actual states of affairs in the outside world: it is because horse do actually have four legs that I see them as having four legs.[54]

[54] Here I am deliberately setting aside possible issues connected with the actual representational import of percepts and eventual counterexamples to the ones I am presenting in these passages. I do not have the space and the required competences to delve into debates in philosophy of perception about these issues. I am rather relying on a naïve but solid

iii) *Already known* from the start by the imaginer. As Sartre famously claimed: "nothing can be learned from an image that is not already known" (Sartre 1948: 12). This means that whatever features the imagining presents, it is there because we put it there. Therefore, I could never learn anything from an imagining as I am the one and only maker of it. I cannot learn from my imagining of a winged horse that a winged horse has four legs if I did not already represent it as having four legs. Here again, we can make a comparison with percepts. If I were to see a horse for the first time in my life – supposing that I did not know *anything* about horses until that very first encounter – my perceptual acquaintance with it can teach me a lot of things about them: e.g. that they have four legs, a tail, etc.

Taken all together, these three features dramatically undermine the possibility for imagination to produce knowledge. However, there are excellent reasons to resist this claim. Hopefully, the exploration of empathy and empathetic imagination I have offered throughout the present work is already enough to raise at least some mild skepticism towards the exclusion of imagination from the realm of knowledge-giving cognitive processes. At the very least, we might say, imagination can give us phenomenal insight into somebody else's perspective. Empathy should by itself be enough to question at least two of the fundamental attributes that have been traditionally attributed to imagination: namely ii) and iii). The imagination deployed by empathy is indeed neither unconstrained nor epistemically sterile. Indeed, empathetic imagination is ideally governed by the epistemic constraint of accuracy. As we should know by now, we aim to *accurately* represent some relevant features of somebody else's first-person perspective in a given situation. The imaginative work of empathy is therefore concerned with the re-creation of certain experiential features that are not instantiated by the actual perspective of the imaginer. Ideally, the imaginings brought about by empathy should be faithful replicas of the other person's actual perspective. If the empathizer were to sneak in the empathetic representation – due to e.g. some cognitive bias – features that are different from those that are relevant for accurately representing the other's perspective, then we would be entitled to claim that the empathic attempt has failed.

As to what concerns iii), imagining the perspective of someone else can provide us with knowledge about what it feels like to be in their shoes. Most importantly, this knowledge is achieved precisely

conception of perception, one that is merely functional to a better exemplification of the features of imagination I am trying to discuss.

thanks to the imaginative representation that we set up. In other words, it is by imaginatively placing ourselves in somebody else's circumstances that we can then assess how this person might feel or what they might be thinking.

It seems then that imaginings can at least lead to the acquisition of some relevant knowledge about other people. One of the examples we deployed in *Chapter II* and mentioned by Lombrozo and Aronwitz (2020) is again useful to illustrate this point. A friend attended a beach wedding in a bathing suit, only to discover the other attendees in formal attire. If we are interested in knowing how this friend might feel in such circumstances, one way to go after the answer is to imaginatively place ourselves in her situation and try to imagine how she might feel, i.e. she is likely to feel embarrassed. Now, the answer to the question "how does she feel?" was not known before we actually placed ourselves in our friend's situation and tried to see what it might feel like to undergo a similar experience.

In any case, it is not the question as to how imagination might lead to knowledge that interests me here, but rather the *answer* provided by Kind to this issue that turns out to be particularly relevant for present purposes. In order to account for the way in which imagination might lead us to knowledge Kind introduces an account of imagination that she calls "imagining under constraints". The basic intuition behind this account is rather simple: our imaginings can be scaffolded alongside some constraints that allow us to structure them in an epistemically productive way. She mentions two major constraints that can guide our imagination:

a) *The reality constraint*.
   Our imagining should conjure up a representation of the world as it is. For instance, if we are trying to imagine how our friend in Aronwitz and Lombrozo's example might feel, we need to imagine the scene as it actually was experienced by our friend. We need to imagine her attending an elegant party and realizing that her outfit does not conform to the party's required dress code.
   It should be noted that for an imagining to respect the reality constraint it is not necessary that it represent *every single* aspect and detail of the state of affairs imagined, but only the relevant aspects of it are imagined. For example, to assess the feelings experienced by our friend attending the elegant party with flip-flops, it does not really matter what colors we imagine the flip-flops to be: whether they are red, orange, or blue. What matters is that we imagine our friend attending an elegant party with an inappropriately casual outfit.

b) *The change constraint*.

Our imaginings should represent the actual causal chain governing the states of affairs and events imagined. Aronwitz and Lombrozo (2020) provide another useful example for our purpose:

> Imagine that you have two cylindrical glasses of the same height, where one is wide and one is thin. Each glass is filled with water to the same height. If you slowly tip both glasses over at the same rate, which glass will spill water first? Or will they both spill water when tipped to the very same angle? This challenge […] is often answered incorrectly. But if people are invited to imagine holding out their hands (as if holding both glasses), to close their eyes, and to mentally rotate their hands until they think the water would spill, they often produce the correct response: that water from the wide glass will spill first. (Aronwitz and Lombrozo 2020: 1)

When people come up with the correct answer, we can claim that they came to learn something in virtue of what they just imagined. They came to know something that was inaccessible before they actually conjured up the relevant imaginings.

What is important for us to notice now is that people actually come to the right answer if they imagine the relevant changes that are produced by the imagined process.[55] It is in virtue of their accurate imaginings of the movements of the differently shaped glasses that they come to know from which one water will spill out first. As already noticed for the reality constraint, for the change constraint as well it is not necessary that the imaginer actually represents *every single* change linked to the states of affairs imagined for the imagining to work in an epistemically productive manner. For example, it is not necessary for us to imagine the relevant neural changes produced in our motor cortex by the movement of rotating the two glasses. These are certainly changes that are linked to the dynamics imagined, but they are *irrelevant* to bring about the answer we are after. Here again, only the *relevant* changes need to be imagined.

Equipped with the above considerations, we can now come back to the concerns raised by Goldie and echoed by Paul and try to at least partly downscale the impact of their criticism against empathy's epistemic possibilities.

---

[55] Sure, people can also come to the right answer by chance, but here – it goes without saying – we are interested in knowledge that is *justified* by the imaginings that produced them. Simply getting at the correct answer by chance or by a lucky mistake is not epistemically justified.

The process of attributing character or personality traits can be seen as a special form of imagining under constraint. Indeed, personality and character traits allow us to scaffold our imaginings in ways that respect both the reality constraint and the change constraint.

First of all, they allow us to resort to a set of generalizable features about a certain class of people. We take these features to actually be instantiated by people belonging to that category, i.e. we take these generalizations to bear some epistemically relevant information about the world as it is. Using Kind's terminology, this is what counts as a reality constraint. In this specific case, they tell us something relevant – and often accurate – about how people instantiating these attributes are.

Second of all, these characterizations provide us with some relevant information about how people belonging to a specific class might *act* to certain contextual cues. That is, they help us structure our imaginings in ways that mirror the psychological and behavioral causal chains that can be instantiated by these people in different contexts. For example, coming back to our kindness example, kind people are taken to have a friendly attitude towards other individuals. Features captured by the "kindness" category inform us about some relevant causal patterns that might obtain in light of certain given circumstances. For instance, the kind person will likely help someone in need, when they happen to meet one. In Kind's terminology, this is what might count as a "change constraint", i.e. a constraint that guides us in structuring the actual or likely unfolding of a certain imagined state of affairs.

Now, how do these considerations relate to Goldie's and Paul's criticism? Kind's account of imagining under constraints help us see how we can come to achieve knowledge and understanding by means of imagination. Moreover, and more relevantly for our present purposes, they provide us with a way out of the epistemic prison constituted by Goldie's criticism. Imagining under constraints is indeed what, in principle, could allow us to achieve understanding about different others. By means of the epistemic constraints under which we scaffold our imaginative exploration of other people's perspectives, we can come to appreciate points of view that do not spontaneously align with our own, i.e. we can re-structure our egocentric map in ways that allows us to escape solipsistic threat highlighted by Goldie and echoed by Paul's considerations of PTEs.

4.4. Explanation-aided understanding

Even if Kind's account of imagining under constraints undoubtedly constitutes a powerful tool that allows us to engage with Goldie's criticism of empathy and Paul's account of PTEs, there are still some reasons not to consider this solution as a definitive weapon against these issues.

Indeed, one might rejoin with skepticism claiming that, even if imaginative constraints can lead us to some knowledge and understanding, they still do not allow us to meet the epistemic requirements of empathy. That is, they do not guarantee the achievement of phenomenal insight.

As already anticipated in the previous section, trait attribution strategies can, at best, provide us with explanations or predictions rather than with an empathetic understanding of other people's perspectives. This means that they are still a useful tool that can productively contribute to improving our overall mindreading skill, while at the same time being of no use for empathy. As stated multiple times throughout the present work, explanations and predictions can still fall short from providing us with empathy.

In light of this, one could be tempted to dismiss Kind's account of imagining under constraints as a persuasive way to spell out imagination's contribution in our quest for knowledge, while at the same time leaving untouched the criticism of empathy we illustrated in the previous sections.

However, there is a way to show how explanations can serve empathetic purposes and thus contribute to our subjective understanding of other people's perspectives. As insightfully noted by Kenneth Kendler and John Campbell (Kendler and Campbell 2014) explanations can work in ways that aid empathic understanding. Their account of *explanation-aided-understanding* states that explanations can ground our imaginative understanding of other people for reasons that are interestingly similar to those we mentioned when illustrating Kind's account of imagining under constraints.

As already seen in *Chapter I*, when I presented the distinction between explanation and imaginative understanding (or empathy) drawn by Jaspers and then reintroduced in the contemporary debate by John Campbell, explanations highlight certain causally relevant correlations among different variables. This strategy allows us to answer satisfactorily to certain kinds of why-questions. With a concrete example, if we wonder *why* our friend Andrew helped the old lady cross the street, we can explain his behavior by means of a trait attribution strategy: Andrew is a kind person.

However, Kendler and Campbell's account of explanation-aided understanding states that explanations can do something more than this, namely they can ground and *scaffold* our (empathetic) understanding. Using our favored terminology, they can enable phenomenal insight.

Specifically, Kendler and Campbell are interested in the way in which neuroscience, by detecting certain kinds of causal chains correlating neural activities with specific thinking and behavioral patterns, enable us to set our imaginative understanding on the right track, so that we can come to

appreciate how these thinking and behavioral patterns might feel from a first-person perspective. According to them, we can "translate" explanations into the language of empathy. They take one of the cases discussed by Jaspers and described by him as un-understandable, i.e. as constituting a case where we experience an epistemic wall out empathic skill cannot overcome. This example is what Jaspers called "delusional atmosphere", i.e. a peculiar psychological pathological state that can be described as "the intrusion of inappropriate meaning into consciousness leading to ideas and delusions of reference" (Kendler & Campbell 2014: 2). Here is how they describe their account of explanation-aided understanding:

> A wide range of research in rodents and primates indicates that midbrain dopamine (DA) neurons not only play a critical role in reward and positive motivation, but also provide signals which ' tag ' other forms of salient experiences including aversive and alerting events. One class of DA neurons reflects the motivational value of stimuli and others, of particular interest to us, encode motivational salience. […] firings of these DA neurons that encode motivational salience provide a signal to the organism saying: 'Pay attention to this stimulus. It is important. Figure out what is going on!' What would we expect to happen if these DA salience neurons fired at inappropriate times? […] [T]he result would be the incongruous intrusion of meaning into consciousness. This meaning – carrying with it a sense of significance – would be associated with whatever the individual was then observing in the environment. It then takes *only a small step* to imagine that an idea of reference could thereby be produced. (Kendler & Campbell 2014: 3, emphasis added).

Thanks to the information provided by neuroscientific explanations we can scaffold our understanding of these peculiar delusional experiences so that it then takes "only a small step" to fuel it with phenomenal insight. We are, so to speak, enabled to construct our imaginings in ways that help us to accurately represent experiences with which we are utterly unfamiliar. Once we are provided with an explanation of how these experiences unfold in the context of one's cognition, we know where to "*look at*" when trying to model an experientially vivid imaginative representation of that experience.

> Given our expanded knowledge of neuroscience, we can convert these experiences of the intrusion of meaning into consciousness from the un-understandable to the understandable. […] In normal life, we have all experienced a misattribution of meaning to an environmental stimulus. For example, you are waiting at an airport for a loved one. You see, in a crowd of passengers, a person in the distance who seems to be her. You jump to your feet and, with a rush of emotion, move toward her to embrace. As you get closer, you realize you were wrong. There was some similarity in body size, gait or clothing that caused you to think this person was your loved one. Your emotional response was misattributed. Through an error in your perceptual system, meaning was placed where it did not belong. We can apply this salience model to a related psychotic phenomenon that Jaspers also considered un-understandable: delusional atmosphere. (3)

Neuroscientific explanations helped us searching for the right kind of experiential analogue that we can imaginatively recast in order to empathize with the experience of delusional atmosphere.

We do not have to buy the neuroscientific account provided by Kendler and Campbell in order to appreciate their argument. It might well turn out that the DA neurons are not responsible for the delusional attribution of meaning to random stimuli and yet, the logic of the explanation-aided understanding strategy would be intact.

Most interestingly for our purposes, this strategy could be extended to the less controversial case of trait attributions. Psychological traits help us *search in the right place* to scaffold our imagination: they constrain it in a way that helps us achieve understanding so that all we need is a "small step" to imagine what it might feel like to first-personally occupy that perspective. When we know that someone is a kind person, then we also know that this person is likely to feel the motivation to help someone in need. So, when trying to imaginatively occupy the kind person's perspective, all we have to do is to look for cases in which we also felt the urge to help someone in need: this would allow us to appreciate what it might feel like to occupy the shoes of the kind person. Stated otherwise, the explanation provided by the trait attribution strategy *aids* our imaginative construction of the other person's subjective perspective.

4.5. Are Goldie's and Paul's concerns solved?

Explanation-aided understanding and imagining under constraints come with a cost. By deploying those epistemic strategies, we lose that spontaneity that Goldie regards as necessary for successful empathetic perspective-taking. This means that we have to engage in a cognitively effortful process that requires us to carefully factor in the interpersonal equation the relevant traits and features that constitute the specificity of the other person's perspective.

However, we have reasons to bypass this specific concern. Indeed, if you place unrealistically high epistemic demands, then it is obvious that they will not be met. As already said, Tietjens-Meyer rightly noted that placing those demands on empathy turns the process into something "almost mystical" that utterly misses the point about how people usually imagine the perspectives of others. Especially when it comes to empathizing with different others, it is obvious that we cannot fully exclude a certain form of "propositional thinking" from the process of empathy for the reasons I illustrated in the previous section.

However, there are good reasons to believe that the spirit of Goldie's concerns is right even if misplaced. We can indeed still hold a cautiously skeptical attitude towards trait attribution strategies as an efficient means towards empathic understanding and phenomenal insight. This is so for at least three reasons: the first one (i) has to do with the fact that, usually, characterizations are coarse-grained, i.e. they do not provide details that might turn out to be crucial for the empathic understanding of another, the second one (ii) has to do with what is known in the empirical literature as the *fundamental attribution error*, the third one (iii) relies on epistemic concerns that are similar that we saw with TEs in general and ETEs in particular: we can still fail to properly appreciate what it is like to inhabit a perspective that is captured by traits and dispositions that instantiate attitudes, mental states, and behavior that we do not usually display.

For what concerns (i), we should notice once more that characterizations are by definition *generalizable*. The features ascribed by characterizations and personality traits can, therefore, turn out to be too coarse-grained to effectively bestow empathetic understanding. Indeed, in order to successfully take someone else's perspective, we often need *detailed* information about the specificity of that person's circumstances and psychological outlook. Coming back to our previous example, we can say that there are many ways to be a kind person. One can behave kindly on the basis of a variety of attitudes ranging from emotional states to more cold-blooded processes such as rational deliberation and cost-benefit analysis. Imagine a case in which our friend Andrew who, while walking to his office, bumps into a hobo begging for alms. He swiftly decides to give him a 10$ bill. Why did he do that? Because – we should know this by now – he is a kind person. But one can further ask: how did Andrew's kindness concretely unfold in the context of his psychology? How did his first-person perspective look like in the process of deliberation that led him to help the hobo? Andrew might have felt compassion for someone who is clearly less lucky than him, or he might have felt distress at contemplating the misery in which the poor man lives, or, he might have cold-bloodedly judged that giving him 10$ was the right thing to do: his finances would have been left utterly untouched and the hobo would have been able to buy a meal.

Which one of these processes took place in Andrew's head? Characterizations, *by themselves*, cannot tell us much to answer that question. We can still be in fail to understand what it actually felt like to occupy Andrew's perspective when he decided to give that hobo a 10$ bill: we cannot "translate" the language of kindness characterizations into an experientially vivid representation of Andrew's first-person perspective. Therefore, character traits, even if they can contribute to scaffolding our empathic understanding of others, cannot by themselves help us get a grasp on what it is like to act or think or feel on the basis of them.

We can now turn to the second problem linked to the attribution of traits and dispositions (ii). Indeed, even if we saw that the attribution of disposition can help us scaffold our mutual understanding in ways that can efficiently sustain empathy, we need to be aware of the intrinsic limits of this routine, one that is widely studied by social psychologists. The problem is known in the specialized literature as the *fundamental attribution error* or as the *correspondence bias*.[56] This phenomenon is usually described as the tendency to overqualify the role of stable psychological dispositions in other people's behavior. In other words: "social perceivers often assume that another's behavior indicates that person's stable qualities" (Fiske & Taylor 2020: 314). There is large evidence showing how we can infer a person's attitudes or character traits surprisingly quickly and with very little effort basing our judgments merely on minimal cues such as facial traits, expressions, and even outfit (Olivola et al. 2014; Todorov et al. 2015). This is how Lee Ross and Richard Nisbett illustrate the point:

> The answer we get both from research evidence and from everyday experience is that people are inveterate dispositionists. They account for past actions and outcomes, and make predictions about future actions and outcomes, in terms of the person – or more specifically, in terms of presumed personality traits or other distinctive and enduring personal dispositions. The evidence […] suggests that people automatically – and unconsciously – provide a dispositional interpretation to behavioural information. (Ross and Nisbett 1991: 90, 120-21, 124)

This tendency bears with it the underestimation of other contextual variables that might have been more prominently responsible for the other person's behavior. For instance, suppose that our friend David was to find himself in an analogous situation to the one we imagined happened to Andrew. David stumbles on a hobo begging for alms. Instead of giving him 10$ – as Andrew did –, David ignores the hobo's request and simply moves on. In this case, attributing a personality trait or a stable psychological disposition such as "cynicism" is what could help us explain and understand his behavior. However, it might well be that David simply was in a hurry, or that he did not have cash at the moment. These contextual variables would give us a drastically different picture of David's perspective in that circumstance and could accurately explain David's behavior without having to resort to any stable disposition of his. Instead of imagining David as being guided by a cold-blooded and cynical outlook towards those who are less fortunate than him, we would imagine David as being moved by the preoccupation of being late or by the regret of not having any cash to give the hobo. Depending on whether we resort to a stable psychological disposition (e.g. cynicism) or to contextual

[56] For a comprehensive overview of the empirical work carried out to investigate this phenomenon, an indispensable starting point is (Fiske and Taylor 2020).

variables (e.g. being late or not having cash), we would end up with two radically different empathic outcomes. Of course, there is no way to solve the issue with a solution that generalizes across the diverse contexts in which we might deploy empathy. Indeed, in some cases, dispositions *do* actually play a crucial role so that attributing them is not only epistemically legitimate but also accurate. In some other cases, contextual variables might offer a better understanding of the situation. In most cases, a mixture of dispositions and contextual variables is what might deliver a better understanding of the situation involved. The fundamental attribution reminds us to hold a sort of guarded skepticism towards trait attribution strategies and makes us aware of how they can, in various context, hinder an accurate empathic grasp of the other person.

For what concerns (iii), it might be useful to say something more about what it means to imagine holding certain attitudes that are taken to be instantiated by characterizations. Indeed, as we saw in the previous sections, in order to be profitably used for empathetic purposes, characterizations need to work as constraints that help us *search for* the right mental attitudes and processes that might be instantiated when certain circumstances occur.

In what follows, I will focus mainly on the example of "desire", but my point can be in principle extended to other kinds of mental attitudes. I do this because the category of "desire" is large enough to encompass a large variety of mental attitudes[57] and at the same time particularly problematic to accommodate within an account of empathy aiming at phenomenal insight.

To have a certain psychological disposition, or character, or personality, entails, to some extent, also to endorse a certain set of values and preferences, or, to hold, implicitly or explicitly, certain *desires* that underlie and guide certain behavioral as well as thinking and emotional patterns. Now, in order to imagine oneself as embodying certain traits that are different from the ones that we actually instantiate in our daily life, we would need to imagine ourselves as holding different kinds of mental states, i.e. also certain kinds of *desires*.

As we saw in the first chapter, when I illustrated the epistemic demands that need to be fulfilled in order to achieve phenomenal insight, empathizing with a desire requires something met from merely imagining *that* a certain desire is held by another agent. We need to be capable of imagining what it is like to hold such a desire from a first-person perspective, what it feels like to inhabit a perspective

---

[57] To some extent, notions such as preferences, values, motives, goals, etc. could be framed, at a psychological level, by resorting to the folk-psychological category of desire.

that is "governed", to a certain degree, by that desire. For example, for a childless person to empathize with a mother or a father, it might be required to imagine not only that the parent holds the desire that their child will thrive and flourish but also how that desire can be so powerful to overcome every self-serving interests that guided their life before the birth of their child. That might be a task that outpowers our epistemic and representational skill. Indeed, even if we (supposing we do not have children) might come to appreciate that parents might be moved by those kinds of desire, we might still fail to appreciate what it is like to live by them, i.e. we might fail to embed that desire within the broader context of our cognitive and behavioral conducts. Being able to empathize with a certain desire, indeed, requires us to be able to at least appreciate what is *desirable* about the object of our desire. Coming to appreciate what is desirable about a certain object of desire amounts, to us being *personally sensitive* to that kind of desire, i.e. being capable of first-personally represent the motivational push exerted by that kind of desire. Let me try to articulate this thought more carefully.

By claiming that in order to empathize with certain desires we need to be able to be personally sensitive to that desire as a desire that we ourselves can imagine holding in certain circumstances, does not amount to claiming that empathizing with a desire necessary leads to *endorse* that peculiar desire. You can indeed grasp the motivational and conative push that certain desires can exert over people holding them, without this necessarily entailing that you yourself come to hold that desire. You can still manage to live by that desire for a variety of reasons: maybe you might hold other – so-called "second-order" – desires that have a greater impact on your conduct.

This can be relatively easily illustrated with an example. Take the desire to smoke a cigarette. Imagine that you never smoked or, alternatively, that you once tried to smoke but found the experience utterly disgusting: it smelled awful, it tasted worse, and if repeated over time it would surely have terrible effects on your physical health. In this case, you would fail to empathize with the desire to smoke a cigarette: not only you are utterly insensitive to the pleasures that it is supposed to bring but you are also aware of its devastating effects on people's health. Now, imagine a different scenario in which you once were a smoker. In this case, you might be perfectly capable to recall the desire you once had to light up and smoke a cigarette: you can perfectly recall the pleasures that it gave you, the psychological relief that was linked to it, etc. All that endows you with the relevant representational resources to grasp what is *desirable* about smoking, i.e. you are sensitive to the desire that motivates people's need to light up a cigarette, even in spite of all the bad consequences that this misconduct can have on one's health and overall wellbeing. However, you quit smoking years ago. This means that even if you are able to first-personally appreciate the desire behind smoking, you still do not *actually* hold it in your life. Other attitudes that you might hold – other beliefs, desires, values,

preferences, etc. – filter and silence the motivational push that is linked to the desire of smoking. Your personal sensitivity towards that desire – i.e. your capability of grasping what kind of features of smoking *warrant* the desire to smoke – turn out to be a precious epistemic resource when it comes to interpersonal understanding and, more specifically, to empathy.

The account I just presented is largely in line with Anscombe's considerations on what she called the "desirability characterization" (Anscombe 1957/2000). She famously claimed that in order to make a desire fully intelligible we need to have a grasp on its desirability characterization, i.e. we need to grasp what kind of 'good' this desire is aiming at. A desire becomes fully intelligible only as far as we can see it as aiming towards certain ends, goals, or goods that we are able to represent as choiceworthy. We need to be particularly attuned to certain features of the desired things that *warrant* the desire about them. In the absence of a desirability characterization, our understanding would simply go astray. In a recent publication, Grimm brilliantly illustrates Anscombe's account:

> Suppose, to adopt one of Anscombe's examples, someone were to claim that his overarching goal in life was to collect saucers of mud. Suppose too that his other beliefs and desires all aligned with this idea, so that they displayed the sort of internal consistency or coherence we found wanting in the last example.
> Anscombe's basic insight is that even though there would be some sense in which we could discern structure in the person's actions (relations of dependence between beliefs, desires, and actions, e.g.) and thus achieve a kind of intelligibility or understanding, in another way the person's actions— guided by the dedication to saucers of mud—would seem extremely foreign and unintelligible. It would be a surd that would simply stare back at us.
> [I]ts [Anscombe's account] implications for understanding human beings are wide-ranging. For it helps us to appreciate that part of what makes an action intelligible for us is to be able to see the goal of the action as not just desired but as desirable, as not just chosen but as choiceworthy.(Grimm 2016: 217).

In our terms, the desirability characterization is what enables us to move from the domain of explanation – i.e. being capable of discerning some *structure* in another person's mental states and behaviors – to the peculiar kind of understanding delivered by empathy – i.e. an appreciation of how these mental states concretely unfold in the context of somebody else's psychology.

It is interesting to notice how Grimm, while endorsing Anscombe's view, is not sure whether this understanding can be achieved by means of empathy (Grimm 2016: 217). However, I cannot see how else it might be possible to make desires fully intelligible, if not by empathizing with them. Let us consider another example discussed by Anscombe:

> Let us suppose some Nazis caught in a trap in which they are sure to be killed. They have a compound full of Jewish children near them. One of them selects a site and starts setting up a mortar. Why this site? Any site with such-and-such characteristics will do, and this has them. Why set up the mortar? It is the best way of killing off

> the Jewish children. Why kill off the Jewish children? It befits a Nazi, if he must die, to spend his last hour exterminating Jews. (I am a Nazi, this is my last hour, here are some Jews.) Here we have arrived at a desirability characterisation which makes an end of the questions ' What for?' (Ancombe 1957/2000: 72)

As Anscombe puts it, getting a desirability characterization is a special way of answering the question 'what for?'. With that question, we are actually enquiring about what features of the desired object make it appear as *choiceworthy*. Now, in the example of the Nazi discussed by Anscombe, we are perfectly capable of seeing what goals are taken to be good by a Nazi. In this sense, we are perfectly able to give a "characterization" of the Nazi's desire. Still, I am sure that most of us would still find the Nazi's desire unintelligible. Even if we know what kinds of goods the Nazi is aiming for and even though we know that certain actions "*befit*" a Nazi, we are uncapable to empathize with them. This means that we are incapable to place ourselves in his situation and imagine what it would be like to have that desire, i.e. what it would feel like to be actually motivated by that kind of desire and committed to the aberrant ends that are taken to be good in the context of a "virtuous" Nazi conduct.

To sum up, finding the link between certain ends that are taken to be good can still be insufficient to bestow empathic understanding of these states. Indeed, I might also realize that the good connected to killing Jewish children can be grounded in the Nazis' pleasure in doing it – pleasure being the good their desires are aiming at. Still, I can fail to appreciate how such a connection might actually be realized in someone's psychology. In other words, I might fail to appreciate the actual motivational – and, perhaps, hedonic – push linked to such desire: I am utterly blind with respect to what it might feel like to hold such a desire. We can then be less wary than Grimm about the link between empathy and the intelligibility of desires and claim that only by placing ourselves in the other person's shoes, we can come to appreciate their desires as desires that we *ourselves* might hold, were we to find ourselves in the specific (internal and external) circumstances our targets find themselves in. By empathically grasping what is desirable about certain goods or ends connected to a certain desire, we can make it utterly intelligible to us. Conversely, it is by failing to empathize with it that we ultimately certify a desire's un-intelligibility.

4.6. Taking stock

Once we appreciate the epistemic demands behind the process of making certain kinds of desires intelligible, we can also come to better appreciate another reason why people can fail to empathize with PTEs.

We saw that one problem with PTEs consisted in the incapability to foresee what kinds of changes in one's psychological outlook would be likely to follow from undergoing a TE. In the previous sections,

I tried to show how the PTEs represent not only a serious threat to our forecasting skills but also a serious epistemic threat that is similar to that posed by ETEs. Indeed, we might imagine a case in which we are perfectly capable to predict *what kind* of changes are likely to occur, if we were to undergo a certain TE and still being incapable to first-personally represent those major psychological changes that the experience might bear with it. This case is easily grasped when instead of considering cases of self-directed predictive empathy – which are the cases that most interest Laurie Paul in her 2014 book –, we consider the case of other-directed retrodictive empathy. In that case, as we saw, we have, in principle access to the relevant psychological change that might be caused by the TE in question. And yet, we might still find ourselves facing the epistemic that makes our empathetic ambitions collapse.

I tried to illuminate this latter point with the help of Goldie's critique of empathy. As we saw, his critique, while overemphasizing issues that are ultimately neglectable – i.e. the "falsification" of the other's perspective by bringing to the foreground elements of their psychology that are embedded in the unconscious background of the target's perspective –, allows us to highlight some radical issues that are likely to emerge when trying to empathize with individuals with psychological profiles that are considerably different from our own. I finally tried to highlight this problem by focusing on the difficulty linked to first-personally imagining desires that we do not usually endorse.

My considerations, hopefully, helped me emphasize even more the epistemic challenge that PTEs pose to empathy. PTEs highlight some genuine limits of our imaginative skills that are not merely linked to the issue of forecasting but rather to the issue of accurately modelling through imagination perspectives and psychological outlooks that considerably differ from the one we usually inhabit.

Onwards

It should be clear by now that the issue of TEs represents a particularly powerful case to highlight some crucial shortcomings of empathetic perspective-taking.

In virtue of the "bounded self" account that I highlighted in the previous chapter, we were capable of seeing how the radical (epistemic and personal) changes that are involved in TEs hinder our empathic attempts because they overflow the representational capacities of empathizers.

However, as I will try to show in the next – and final – chapter, the idea of a "bounded self" does not in principle preclude any possibility to move beyond the narrow perimeter of our egocentric perspective. It actually leaves space for some interesting and surprising epistemic and personal gains

that can be achieved through empathy. These are ultimately linked to our capability to *stretch* our representational capability in ways that might either expand our representational capacities or even change the kind of persons we are.[58]

All that will keep us busy in the next chapter.

[58] The disjunction is inclusive.

Chapter V

What Empathy Can Do: Learning Trough Empathy, Changing Through Empathy.
The Epistemic and Personally Transformative Power of Empathy

1. Introduction

In the present chapter, I am going to see how the challenges to empathy explored in the previous chapters can be counterbalanced. This chapter should not be taken as being in contradiction with the previous one. The arguments that I am going to present in the following sections and those offered in the previous ones are not mutually exclusive. Indeed, rather than trying to show how Paul's and Goldie's challenges to empathy can be completely overcome, I will try to show how they can be *faced* by empathizers. This will allow me to show some surprising epistemic features of empathy that are seldom discussed in the contemporary literature.

I will deal with two underexplored outcomes that empathic processes can have. These two outcomes can be modelled alongside Paul's discussion of TEs. As we will see, there can be a strictly *epistemic* outcome and a broadly *personal* outcome that can stem from the process of empathy. I will very briefly anticipate what both issues are about before delving into them more in detail in the following sections:

i) *Epistemic outcome*. The epistemic outcome I have in mind is one that has to do with empathy's capability of giving us an insight into experiences that we did not personally undergo in our life. That means that empathy, in some cases, can make us achieve phenomenal insight about experiences we never had.

ii) *Personal outcome*. The personal outcome refers to the personal changes that might occur as a consequence of engaging in empathic processes. The basic idea is fairly simple: by making us acquainted with perspectives and points of view that are different from those we usually occupy, empathy brings with it the possibility of restructuring our set of preferences, goals, values, desires. By empathically realizing that there can be a different way of looking at the world, we might come to actually endorse that new perspective.

Another way in which it is possible to summarize the previous considerations is by utilizing Paul's influential terminology. Empathy itself can indeed count as a *transformative experience*, i.e. it can entail an epistemic and a personal transformation. By teaching us something new about – for us – inedited experiential states it can be said to be *epistemically transformative* and by making us acquainted with different viewpoints, it can change who we are, i.e. it can be *personally transformative*.

2. Moving beyond the black-and-white model of epistemically transformative experiences

In the previous chapter, we illustrated the nature of ETEs by referring to examples that had an intuitive appeal, i.e. it was easy to extrapolate the lesson that they were meant to convey. For example, in the case of Mary, it was immediately evident how she could have no clue about the experience of seeing colors so that her first acquaintance with the experience of seeing colors clearly seemed to endow her with some knowledge that she could not have accessed before. Similarly, Paul's thought example of becoming a vampire succeeds in pointing at the epistemic wall that impedes us to accurately imagine what it would be like to be a vampire, i.e. what it would be like to occupy a point of view (or a perspective) which involves a radically different way to experience the world.

However, as already anticipated in the previous chapter, what makes TEs so fascinating is that they are not limited to the mere case of thought experiments. According to Paul, our lives are recurrently shaped by these kinds of experiences.

Let us consider again the experience of becoming parents (Paul 2014, 2015). Paul stresses how parents seem to be acquainted with experiences that are different from any other they had before (epistemically transformative) and, at the same time, seem to go through radical changes as individuals (personally transformative): their values, preferences, desires, and beliefs evolve in a different, unforeseeable way.

Given all that, our intuitions might convey this kind of judgment: it seems impossible for people who are not parents to empathize with the (transformative) experience of becoming parents.

> Now, having a child is not just a radically new experience. For many people, it is also a life-changing experience. It might be wonderful, or joyous, or happy—or it might not. But however it is, it is usually very intense, and people who have a child and respond in the normal way find themselves with very different perspectives and preferences after the child is born. That is, for most people, having a child is an epistemically transformative experience that is also personally transformative. Your preferences will change. The way you live your life will change. What and who you care about will change. (Paul 2014: 81)

And she concludes:

> I conclude that having your first child, in many ways, is like becoming a vampire. (82)

The question to ask now is whether it is fully appropriate to conceive of real-life occurrences of transformations alongside the thought experiments such as Jackson's Mary and Paul's vampires? Paul clearly states that real-life TEs are "structurally parallel to […] Frank Jackson's case of Mary growing up in a black and white room" (Paul 2015: 764).

I believe that, even if they do a fine job in illustrating what both epistemic and personal transformations amount to, they shape our intuitions in a way that impedes us to appreciate some crucial details about what is actually going on in the case of real-life TEs.

This way of conceiving of TEs has been recently dubbed by Jenann Ismael (Ismael 2019) as the "Black-and-White Mary Model". In such a model of TEs, the epistemic obstacle is clearly insurmountable. As we saw for Mary, she utterly lacked access to the phenomenology of color visions simply because she never was personally acquainted with it as she spent her entire life in a black and white room. However, many of our experiences present a much more complex and multifaceted profile than the imaginary experiences mentioned in Mary's story and in the vampire-case.

The issue of the Black-and-White Mary model allows us to appreciate how confronting radically different perspectives is something that is not perfectly captured by those clear-cut transformations mentioned in the above examples. The radically different perspectives we happen to encounter in our usual perspective-taking routines do not involve such well-rounded transformations whose epistemic and personal profiles are readily individuated and isolated as absolute new kinds with no direct link to anything else we have been acquainted with. To put it with a slogan: many TEs are much more a matter of degree than a matter of pure kind. In many cases, it is much more about the uniqueness of their "gestaltic" constitution rather than the phenomenological uniqueness of their constituents. Stated otherwise: the uniqueness of becoming parents is not necessarily due to the epistemic uniqueness of its experiential constituents, but rather due to the uniqueness of their combinatory array. Therefore, the constructive and combinatory nature of empathic imagination could in principle succeed in getting a grasp on – at least – some TEs.

If that sounded rather cryptic, I will clarify everything in what follows.

3. A closer look to the parenting example: from an experiential "gestalt" to an experiential "compound"

Let us take a closer look at the parenting example discussed by Paul so that we can come to appreciate more concretely the nature of the considerations introduced in the previous section.

The experience of having a child is usually said to be accompanied by an intense, distinctive, and multifaceted phenomenology. The very first worry one might have when confronted with such a case is that the experience of being a parent can cover a remarkably extended period of time. Indeed, there is no single experience that compounds together all – or at least most of – the essential features of the experience of being parents. Usually, people not only *become* parents but also *remain* parents for a long time: parenting is, therefore, better conceived of as a *cluster* of different experiences that can be grouped together under a large experiential category. For instance, parents might undergo the experiences of joy and astonishment when holding their newborn for the first time, go through the complex process of becoming emotionally and personally attached to their children, experience a dramatic shift in their set of preferences or values. All this – and much more – might come into play in shaping the multi-layered experience of becoming parents.

Therefore, the first issue one faces when trying to empathize with the experience of being a parent is that it involves not a single experience but *many* experiences, usually extended over time, most of which come with a complex and distinctive phenomenology.

This worry has been recently – and very briefly – addressed by Kendal Walton (Walton 2015) who introduced the term *situational empathy,* which was suggested to him by Jonathan Weinberg.[59] Here is what Walton – paraphrasing Weinberg's words – has to say about it:

> Situational empathy, he [Jonathan Weinberg] says, concerns "phenomena that are extended in time, often involving distinct qualitative characters at different times"; it "includes not just the phenomenology, but affective, practical, behavioral, cognitive, etc. elements". […] Situational empathy comes in many varieties, and there is a lot to be said about it. (Walton 2015: 15).

Unfortunately, even if there is "a lot to be said about it", Walton does not tell us much about it. I can try to say something more.

---

[59] Kendal Walton informs us that the expression was suggested to him by Jonathan Weinberg in some comments he made while commenting on a draft of his paper (Walton 2015: 15)

The experience of being parents nicely fits the description of situational empathy: it is extended over time, involves distinct qualitative features, and has huge behavioral and cognitive implications for the people undergoing it. Since there is no single experience that holds the exclusive right to be called "being a parent", the first thing an empathizer can do is to break down this broad experiential category into smaller experiential units, i.e. the empathizer needs to have a narrower and epistemically more realistic goal when trying to achieve phenomenal insight about the parenting experience. Indeed, I suspect it is much easier to perceive the epistemic wall represented by the experience of being parents as insurmountable if we approach it as a gestaltic unit with a unique phenomenal character that cannot be reduced to anything else we experienced in the past. Once we see that this massive experiential category can be de-composed into smaller experiential units, then the empathic work starts looking slightly less hopeless.

If you look at the experience as an indistinct whole, you might well have the impression that insight into those kinds of experiences is utterly impossible. But if you start looking at them as experiential compounds that can be imaginatively broken down by focusing on smaller experiential bits to which we can relate to in some way, then there is a chance for empathy to succeed.

In a nutshell, the first move that we can do when addressing central cases of TEs such as the parenting example, instead of looking at them as an *experiential gestalt* we might instead look at them as an *experiential compound*. An experiential gestalt is a clearly defined experiential kind that cannot be decomposed into smaller units and constitutes a homogenous phenomenological category, an experiential compound is a multilayered array of experiential states that can be grouped together to form a distinct experiential category and at the same time be distinguished from one another. In other words, an experiential gestalt can be conceived of as an experiential state that cannot be related to anything else without losing some fundamental kind-defining property. An experiential compound is instead a complex experiential state that finds its identity not much in the irreducible uniqueness of its phenomenological profile but rather in the combinatory array of multiple experiential states.

Looking at certain instances of TEs as experiential compounds rather than irreducible experiential gestalts might be a promising way to gain at least a partial insight into what it feels like to undergo them. I believe that this way of approaching the issue of TEs allows us to take into account the concerns raised by Ismael against the Black-and-White Mary model. I, indeed, share Ismael's concerns about limiting our analysis of TEs to this model would make us blind to more interesting cases where rather than with "black-and-white" experiential gestalts, we have to do with "grey"

experiential compounds, which certainly challenge our empathic ability while leaving some space for interesting epistemic discoveries.

Before we see how this process might look like more in detail, it is useful to stop for a moment and see the kind of criticism that is usually directed to empathy as a means to achieve an insight into radically different perspectives. This criticism is informed by a large empirical literature in social psychology which succeeded in showing how bad people usually are at taking other people's perspectives.

4. Troubles with perspective-taking

A wealth of empirical research has successfully illustrated the mechanisms behind perspective-taking processes. Works in social psychology and neuroscience have shown both the epistemic and broadly moral consequences of taking the perspective of someone else.

On the one hand, experiments in social psychology and neuroscience has shown how perspective-taking can be a valuable tool to reduce stereotype attribution and inter-group bias (Todd et al. 2011), and that it can have a remarkable positive impact on our altruistic and generally prosocial dispositions. Existing work has demonstrated our ability to rapidly take the other person's perspective (Samson et al. 2010), and there is work showing our tendencies to mimic or imitate other people, without our being conscious of it (Chartrand and Bargh 1999), and even when doing so is counterproductive to our own goals (Naber et al. 2013).

On the other hand, there is notable empirical literature showing the obtrusion of our egocentric point of view when trying to figure out what other people are thinking and feeling.[60] In the present section I am going to address some relevant work that has raised if not some outspoken skepticism it has at least offered some good reason for caution towards perspective-taking mechanisms.

A caveat is in order before we can move on. The literature I am referring to deploys several terms to describe what I have called "empathy". Psychologists and behavioral scientists have used terms such as empathic projection, simulation (Shanton and Goldman 2010), and social projection (Krueger and

---

[60] This is a concern that I already addressed in *chapter III* to make my point about the "bounded self". In this section, I am going to refer to this work with a different point in mind: articulating the reasons behind the scepticism against perspective-taking that has been fuelled by relevant psychological work on the topic.

Clement 1997). It goes without saying that these different threads of research focus on different aspects of the general process of taking somebody else's perspective. Furthermore, it is hard to find a study whose operationalized notion of perspective-taking perfectly matches what I have been calling empathy so far. However, this should not prevent us to confront with those results. As I already stated in *chapter III*, even if the notions of perspective-taking deployed in these works do not perfectly coincide with the account of empathy I developed in the present work, the results they produced can still be taken into account and fruitfully discussed. As we will see, some of the concerns related to generic processes of perspective-taking highlighted by this research can be easily related to the narrower and philosophically more detailed account of empathy of the present work.

4.1. Egocentric bias

The first big issue we need to tackle has to do people's reluctance in getting over their own perspective. Our own egocentric perspective can indeed dramatically impact how we understand others, often compromising the accuracy of our intersubjective assessments (L. V. Boven and Loewenstein 2003; Epley et al. 2004; Epley 2008; Eyal et al. 2018; Loewenstein and Boven 2000). This is what is usually called in the specialized literature *egocentric bias*.

In study experimenters (Keysar et al. 2003) showed some interesting shortcomings of adult subjects in practically discriminating between their own beliefs and another subject's beliefs. In one of the experiments, a "director" instructed a participant to move certain objects around in a grid. Participants had previously hidden an object (i.e. a roll of tape) in a bag. They knew what was in the bag, but they also knew that the director did not know. Nevertheless, when the director asked to move an object whose description matched not only the identity of the object that was mutually visible, but also the identity of the hidden object (i.e. "move the tape", where "tape" could both refer to the hidden roll of tape or to the mutually visible cassette tape) they often moved the object in the bag.[61] This is clearly an example of projection of one's own beliefs onto the target and this is an eloquent case of what Goldman called "quarantining failure" (2006), i.e. the mindreader's failure to track a mismatch between his own and his target's mental states in a way that prevents unwarranted projection of them onto the target.

In another study (Van Boven and Loewenstein 2003) participants had to answer some questions either before or immediately after having exercised. Experimenters believed that this variable would

---

[61] More precisely, 71 % of participants tried to move the object in at least one of the four cases where there could be a misidentification between the hidden object and the mutually visible one.

influence their feelings of thirst and warmth. Participants read about three hikers lost in the mountain without any food or water and were asked to answer whether hunger or thirst would be more unpleasant for the hikers. As expected, differently from the participants who did not exercise, people who just trained picked more often water over food.

In a series of experiments run by Savitsky and colleagues (Savitsky et al. 2001) participants who were exposed to embarrassing situations (actors) tended to estimate that the participants observing them (observers) would judge them more harshly than they actually did. This is not a case of simple mental state projection – actors did not merely project their embarrassment onto the observers. Still, it is clearly a case of egocentrically biased mindreading: indeed, their evaluation was greatly influenced and epistemically crippled by their actual state of embarrassment which impaired a more accurate evaluation of the observer's reactions. This is also known among psychologists as the *spotlight effect,* i.e. people's tendency to overestimate the salience of their actions in the eyes of others in virtue of the mere fact that it is salient from their own egocentric point of view (e.g. Gilovich et al. 2002).

In another study (Galinsky and Moskowitz 2000), participants were shown with photographs of people and asked to put themselves in their shoes and imagine how their days would look like. It turned out that people were engaging in a mixture of egocentric projection and stereotypical characterization.

Mismatches between empathizer's and target's mental states in a given situation have to be carefully taken into account for perspective-taking to be accurate: the mindreader needs to quarantine those mental states that have "no counterpart in the target" (Goldman, 2006: 148) in order to get them correctly.

Interestingly, egocentrism is an elusive issue that exerts a pervasive influence that is not limited to the mere projections of some of one's own specific mental states onto the other. Our egocentric perspective, in fact, functions as a lens of pre-existing beliefs, emotional dispositions, attitudes, moods, intentions, etc. that can result in a remarkable mischaracterization of our target's perspective. This is what has been called by Nicholas Epley (Epley 2008) the *construal bias:* our egocentric perspective alters the way in which we interpret and encode the perspective of other people. This means that gaps do not merely come as punctual and isolated mismatches between empathizer's and target's mental states, but also at a broader level of abstraction that includes psychological dispositions, preferences, etc.

It is easy to see how this tendency to unwarrantedly project our own traits onto the other has fuelled some discontent with perspective-taking: how can we ever aim at successfully taking somebody else's perspective if we have so many troubles getting over our own?

4.2. Affective forecasting

Studies on the so-called phenomenon of *affective forecasting* have also offered a privileged angle to highlight some major shortcomings with our capacity to imaginatively project ourselves into perspectives that are different from the one we are actually occupying. Affective forecasting refers to our ability to anticipate the hedonic import of certain events by allowing us to pre-experience them via simulations in our minds.

The fact that humans are capable of episodically imagining the consequences of experiences without having to personally undergo them is certainly one of the most remarkable miracles of evolution, one that has enabled creatures like us to avoid pains and pursue pleasures before we can actually experience them. However, our ability to accurately simulate the hedonic quality of future events has been put into question by leading scholars in the field arguing that simulations are epistemically crippled in virtue of some of their very essential features.

Daniel Gilbert and Timothy Wilson (D. Gilbert and Wilson 2007) noted how people's hedonic responses can be attributed to both to our *mental representations* of the event and to the influence of *contextual factors* on our simulations. The logic at the basis of affective forecasting's failures is fairly simple to grasp: if we are interested in understanding our hedonic response to some event in the future, we need to be able to construct mental simulations that are accurate enough to work as a proxy to what is actually going to happen. This means that my mental representation at time T1 – at the time in which I run the imaginative simulation – needs to be similar enough to my mental representation at T2 – at the time in which I actually undergo the experience – and that the contextual factors C1 at T1 exert a similar influence to that exerted by contextual factors C2 at time T2. Stated in other words, the simulation and the actual "perception" of the event in question might coincide – to a reasonable extent at least –, and the contextual influences on both representations should be the same. However, so the authors argue, very often at least one of these conditions is not met. Our simulations are indeed subject to representational biases of various kinds that push our prospections far from the standards of accuracy we might desire from them.

On the basis of these considerations, Gilbert and Wilson enlisted four major issues with simulation. I will address them in turn:

i) *Simulations are unrepresentative*.

We construct simulations of future events by relying on the memories we collected of our past experiences.[62] This simply means that the memories we imaginatively "recast" to run the simulation should fit the representational structure of the event we are trying to foresee. However, research has shown how often we rely on unrepresentative memories in order to simulate the impact of a similar future event.

People, indeed, tend to remember much more vividly unusual events because they are so memorable. The tendency to better remember *memorable* – indeed – experiences might have epistemically counterproductive effects on the way in which we simulate future events. Memorable experiences are not only memorable but also – alas! – also unrepresentative of our usual experiences. And that is precisely why they are so extraordinary. But, if we are going to rely on memorable experiences to simulate a future experience, then of course our simulation would likely be inaccurate because it would be constructed on the basis of an unrepresentative model. For instance, I can perfectly remember that day in which at a coffee shop in Paris I met my future partner, but I have basically no memories of that time in which I went to the same coffee shop one week before meeting my partner and nothing worth mentioning happened. If I build my representation of my future visit to a coffee shop in Paris on the basis of my extraordinary memory, then I would likely overestimate the hedonic impact of my future experience. In one experiment, people waiting in line at the subway station were asked to imagine how they would feel if they missed their train that day (Morewedge et al. 2005). It turned out that people were relying on the memory of some of their worst train missing experiences. As a result, they overestimated how painful missing the train would be for them. This is because we can much more easily remember that time in which we missed our train and lost the job of our life, rather than that time in which we missed the train and took the next one two minutes later.

ii) *Simulations are essentialized*.

When we imagine an episode, we tend to reduce its representational structure to its essential features. For example, when I try to imagine how my talk scheduled for next week is going to go, I am going to imagine only certain representative features of it: the

[62] I am aware this might sound too vague, but I leave it as it stands for the moment because I will come back to this issue more in detail later in the chapter.

tone of my voice, the contents of my talk, the joke I intend to tell, etc. I am probably not going to imagine some unessential details of the scene such as the hairstyle of the woman sitting in the first row. The problem with that is that, often, what we would deem as unessential details are likely to play an important role in the concrete unfolding of the future event. For instance, imagine that the woman in the first row turns out to have an overly eccentric hairstyle which is going to completely capture my attention throughout the talk. As a consequence of this, I might fail to focus on my talk because I would keep going back to the woman's hairstyle. My talk would turn out to be a disaster because of what I would likely deem as unessential to an accurate representation of my public talk.

iii) *Simulations are abbreviated*.

This aspect is closely related to the previous one. Indeed, we do not only tend to *essentialize* simulations of events, we also tend to *abbreviate* them. The reason is straightforward: if we were to imagine each and every aspect of an event, then the simulation process would last as long as the very experience of the event, making simulations an impracticable method of efficient decision-making. Therefore, we select specific representative moments of the event in question, which we deem as relevant to evaluate its impact on us. This abbreviation process comes with a trade-off which negatively impacts imagination's epistemic reliability. Indeed, as Gilbert and Wilson notice, the hedonic impact of certain experiences usually dissipates over time. If we abbreviate simulations, then we might overlook the hedonic mitigation that is likely to occur as a consequence of time passing.

iv) *Simulations are decontextualized*.

The contextual factors that influence us in the moment in which we simulate are different from those that influence us when undergoing the actual experience. As mentioned earlier, C1 and C2 must have the same influence on our mental representation for it to be accurate. This case can be easily illustrated by the experiment involving the three hikers vignette (Van Boven & Loewenstein 2003) I mentioned earlier to illustrate the egocentric bias. As already stated in the previous section, the affective states participants were in influenced the way in which they predicted how the three hikers would feel. This phenomenon is known in the literature under the label of "hot-cold empathy gaps". The notion is meant to capture people's struggle to accurately assessing the impact of emotional states on their own and other people's behaviors and mental states (Van Boven et al. 2013; Loewenstein

and Van Boven 2000). In the case of "cold-to-hot empathy gaps" people who are in a "cold" state – viz. when they are not affectively aroused – tend to dramatically underestimate the behavioral and psychological effects of an affective – "hot" – state they or other people might undergo in the future. Conversely, with "hot-to-cold empathy gaps", people who are affectively aroused do not seem to realize that the state they are undergoing is biasing their evaluations and that they would probably think differently if they were to make their evaluation after their transient affective drive has "cooled-off". All that can be seen as an asymmetric influence of contextual factors on our simulations and experiences: people often fail to appreciate i) that certain contextual factors dramatically influence the way in which they run the simulation of a future event and ii) that these contextual factors are unlikely to exert the same influence in the future. In sum, on the one hand contextual elements that influence people at T1 (when they run the simulation) can be absent at T2 (when they actually undergo the experience) and, on the other hand, that contextual elements present at T2 can be absent at T1.

These concerns related to affective forecasting have made psychologists wary of simulations and in general of the capacity of imagination to provide us with an actual insight into experiential states other than the one we are presently undergoing.

In light of all this, it is easy to see how these issues related to imaginative forecasts about our own affective states in the future easily translates into issues with other-directed perspective taking: if we are not capable to accurately imagine *our own* future perspective what hope do we have to accurately imagine someone else's perspective?

4.3. Getting perspectives rather than taking perspectives?

The issues related to perspective-taking that I tried to briefly illustrate in the sections above have induced psychologists not only to grow more and more skeptic about the epistemic powers of perspective-taking but also to search for possible alternatives that might be more effective in providing a real insight into other perspectives.

One of the most interesting attempts in contemporary literature comes from a line of research that contrasted the process of perspective-taking with that of perspective-getting (Eyal et al. 2018). In this large experiment, people were tested across a wide range of tasks involving accuracy such as predicting another person's emotions from facial expressions and body postures, predicting fake versus genuine smiles, predicting when a person is lying or telling the truth, and predicting a spouse's

activity preferences and consumer attitudes. These tasks were all meant to measure accuracy in interpersonal insight, i.e. in our capacity to grasp the goings-on in other people's minds.

This experiment turned out to have rather striking results: it (supposedly) subverted the idea of perspective-taking as a reliable mindreading practice. Indeed, when people were asked to *take* other people's perspective, they did not show any improvement in accuracy but rather, if anything, a decrease in their interpersonal insight. The only remarkable increase that the authors were able to track was participants' *confidence* in the accuracy judgments. Overall, perspective-taking did decrease egocentric bias and stereotype attribution – a result that is largely in line with previous research conducted on this topic – but that did not automatically translate into more accurate judgments about other people's thoughts, feelings, attitudes, preferences, and other mental states.

Interestingly, the authors tested another condition, involving what they called "perspective getting" as opposed to perspective-taking. In this condition, rather than being asked to take the perspective of another participant, they discussed with them, so that they could directly *get* their perspective through conversation rather than *take* it via imagination. It turned out that people in the perspective getting condition were much more accurate than people both in the control condition (in which they received no explicit instruction) and in the perspective-taking condition.

These are interesting results that considerably contribute to the general skepticism against the epistemic reliability of perspective-taking. For instance, in one condition of the experiment, participants spent 5 minutes trying to imagine the perspective of their partner. This is a remarkably long time for a task like this. We usually do not spend that much time trying to imagine somebody else's perspective. If not even such a long time improves our interpersonal insight, then there might be some serious issue with the strategy itself.

However, there is at least one good reason to partly resist the conclusions drawn by the authors of this study. They seem to rely on a highly questionable assumption, namely one that takes perspective-taking as imaginary gameplay that we do in isolation, joined by the sole company of our imaginings.

In the perspective-taking condition people read instructions like the following one:

> While watching the pictures [or videos], please think about the person in the picture [or video]. Try to adopt the perspective of the person in the picture [or video] as if you were the person who is answering the question. Do your best to adopt his or her perspective, putting yourself into the other person's shoes as if you were that person. Remember that the person in the picture [or video] may have a different perspective than you do as the viewer of the picture [or video]. (Eyal et al. 2018: 552)

Participants were alone watching a picture (or a video) and could not interact with the person recorded (or portraited). In other words, they were left alone with a body of information that could neither be updated nor revised in light of new evidence coming from the target.

When you move from such an assumption, then it is easy to contrast perspective-taking with perspective-getting, the latter being a more open-minded, interactive, and receptive way to understand other people by taking into account what *they* have to tell about themselves.

However, I believe that only by deploying this rather artificial experimental framework we come to such a clear-cut conclusion. Indeed, if we look closer at how processes of perspective-taking are likely to actually unfold in everyday experience, then it seems that the perspective-taking/perspective-giving opposition is a false dichotomy.

In the perspective-getting condition, participants read the following instructions:

> Before you rate the extent to which your partner would agree or disagree with the following statements, we would like for you to ask your partner to tell you about their opinions. We will give you a list of statements. Please take approximately five minutes to ask your partner about the extent to which they agree or disagree with each of the statements on the list, trying to get a sense of the range of your partner's opinions. Your partner might strongly agree with some statements, somewhat agree with others, and they may strongly disagree with others. Once you have done that, we would like for you to use the information you got from your partner to predict the extent to which your partner would agree or disagree with these statements. Please use this strategy even if you think another strategy would be better. (564)

In this condition, instead of playing alone with their imagination for 5 minutes (as it happened in the perspective-taking condition), participants were prompted to *interact* with their targets in a way that could lead them to gain new insight into their perspectives. It turned out that this process led to a significant increase in participants' accuracy.

Are we sure that, in light of these instructions, we can conclude that the perspective-getting condition represents an option that can be contrasted with the perspective-taking condition? I believe this is far from obvious.

To intuitively illustrate my concern, an example might turn out to be useful. Imagine that we are at a friend's place and she is telling us about a terrible fight she had with her father just the other day. She tells us how the fight started and how it immediately escalated so that she left yelling at her father that she would never talk to him again. As the good friends we are, we start imagining what it might have felt like to be in our friend's shoes. Maybe we also have a conflicted relationship with one of

our parents and can relate to the frustration and pain that might result from such fights. We start to empathize with the range of feelings we can distill from her story, her bodily gestures, tone of voice, and from the stock of memories of similar experiences we had. Then, after a brief break, she unexpectedly reveals to us that, all things considered, she now feels relieved. The relationship with her father was toxic for her and was negatively impacting her general emotional well-being. A cut in the relationship with her father brought with it the promise of relief. We were not expecting such a twist in our friend's story: we were assuming that she might be feeling excruciating pain and anxiety. Clearly, we wrong, so we revise our interpretation.

Enough drama for now. The question we need to ask in light of this story is the following: do we stop perspective-taking after we realize that we were making mistakes in the way in which we modelled the emotional character of our friend's perspective? I believe we can safely answer with a no. We do not stop perspective-taking in favor of perspective-getting, rather, we re-orient our perspective-taking in light of the new information we are able to *get* from our friend's testimony.

The fact that we can *get* new information about someone's perspective by interacting with and talking to them does not imply that we also stop imagining (or *taking*) their perspectives. People's testimonies can instead offer an invaluably effective way to gain actual insight into their lived experiences.

Perspective-taking – and, therefore, empathy as well – allows for adjustments and updating in light of new evidence. We do not stop to perspective-take if we realize we were wrong the first time, we do not stop to perspective-take if someone *tells* us how they feel. Rather, we are capable of sensibly re-orient our imaginings in a way that adheres to the information we just received.

There seems to be a fairly widespread misconception in both the psychological and philosophical domains about how empathy and perspective-taking concretely happen. The main target of scholars' attention has been the processes that happen in the mind of the mindreader and rarely how these processes can be meaningfully informed by the target's actual contribution. However, as insightfully noted by Richard Moran, understanding persons, as opposed to understanding objects, needs to be open to what these persons have to say about their experiences:

> The tomato, the planet and the proton don't have a view on how they are described; there is no issue there for our best theories of these things leaving out some aspect of their existence that matters terribly to them; and our explanations of their behaviour do not wait upon their acknowledgement that, "Yes, this must be what I've been doing all along". (Moran 2011: 252)

And that therefore:

[Persons] *do* have some say in the matter as to what shall count as being known and being understood. (253)

Coming back to the experiment from which we started, Eyal and colleagues were capable to highlight the somewhat oddly neglected aspect of people's "say" on their own experiences and how beneficial it may be for interpersonal insight. Their "mistake" consists, I believe, in contrasting this process with perspective-taking. This results in an artificial distinction between two processes that should not be regarded as antagonistic but rather as complementary. This false dichotomy reinforces the inaccurate assumption that empathy and perspective-taking are encapsulated processes with which one engages in solitude.

As I tried to show, we can conceive of empath and perspective-taking as being open to revision and to integration in light of new evidence. This might sound like a rather simple observation but is crucially important in order to address some of the concerns voiced by psychologists against perspective-taking. If we conceive of these processes as epistemically revisable, then issues such as egocentric bias and affective forecasting, serious and pervasive as they may be, do not represent a knock-down argument for perspective-taking and empathy. They highlight serious flaws in those processes that might nevertheless productively overcome through epistemic recalibration.

Moreover, this process of getting-to-know-more that grounds empathy's capacity to provide an insight into novel kinds of experienced we did not previously have. The amenability of imagination to be integrated with and guided by novel information we manage to gather about certain states of affairs is what allows it to be *epistemically generative*, i.e. capable of providing us with knowledge that was not available to us before we actually engaged in the imaginative process.

I will explain more in detail how this might work in what follows.

5. Gaining insights into new experiences

There is a very obvious sense in which we can say that we can imagine experiences we never had before. When I imagine what it would be like to go to a friend's party tonight, I'm surely imagining an experience that I did not have before. After all, the party is yet to happen.

However, the issue I will be interested in in the following sections has to do with a more radical sense than the one captured by this simple phenomenon of prospection. The kind of novelty I am interested in is the better captured by Paul's notion of epistemically transformative experience. I will argue that we can, in some cases, achieve insight about what it might feel like to have epistemically transformative experiences. And we can achieve this goal through imagination and, specifically,

through empathy. In this sense, the very process of empathy can be said to be epistemically transformative or, as I will also put it, epistemically generative. Let me clarify how.

First of all, we need to identify what the epistemic requirements for successful phenomenal insight might be. Thanks to our discussion of TEs, we can safely state that the epistemic ground that allows empathizers to bridge the *gap* between them and their targets is represented by what we might call their *experiential competence*. By this formula I mean that one's own set of past experiences can fill their imagination with phenomenal properties of a specific experiential occurrence. If we have once experienced the taste of coconut, I are endowed with the representational capabilities that allow me, among other things, to imagine what it feels like to taste coconut. Of course, the necessary precondition for such an experience to work as experiential competence is that we retain a representation of it in our memory. It is rather obvious that if we had a specific experience we did not retain any traces of in our memory, we will not be able to add that experience to the set of experiential competence. An example here might help: suppose I happened to have chickenpox when I was 3 years old. Chickenpox then might surely figure as an experience that I personally underwent. As an undeniable proof for it stands the fact that I am now immune to chickenpox. And yet, since I was so young, I cannot recollect any aspect of that experience. As a result, I am not in the position to accurately represent what it feels like to have chickenpox.

Suppose now I contracted chickenpox at the ripe old age of twenty-seven. Given the absence of any pathological or other contextual impairments to my memory system, I retain at present very vivid memories of that disease. This endows me with the fine-grained epistemic capacities to accurately represent what chickenpox feels like: the unbearable itchiness, the maddening frustration of not being allowed to scratch yourself, etc. This retained information can now serve as experiential competence and enrich my empathetic capabilities.

On the basis of such an example, we can draw the provisional conclusion that, in order to gain phenomenal insight about another person's experience, we need to have the relevant experiential competence, viz. we need to have undergone experiences that are comparable to the one our target is undergoing, and to be able to summon them up. The similarity between our target's experience and our past experience of the same type allows our imagination to be filled with the adequate phenomenology of a certain experiential occurrence.

As intuitive as these considerations may sound, they are still far from being uncontroversial. One might in fact raise issues similar to those highlighted in the previous sections and ask *how similar* the target's and the empathizer's experiences need to be in order for the latter to legitimately aim at

empathic accuracy: is it necessary for the empathizer to have undergone exactly *the same kind* of experience? Coming back to our chickenpox example: does the empathizer really need to have contracted the same disease or can she work out what it feels like to have chickenpox without having contracted it herself? Indeed, she might surely have experienced some occurrences of itchiness and some others of frustration even though they maybe did not happen both at the same time and for the same reason as in the case of chickenpox. In this latter case, it seems that there would still be space for the empathizer to understand what it feels like to have such a disease even though she did not experience it as such.

As we saw when discussing the case of experiential compounds as opposed to pure experiential gestalts, this means that the empathizer can come to learn something new not by experiencing it. Empathy might, therefore, *expand* the epistemic scope of our imagination.

5.1.Phenomenal imagination, phenomenal properties, phenomenal concepts

In a recent paper, Antonia Peacocke (Peacocke 2020) investigated some issues that are in line with those I am analyzing here. The guiding question of her paper is the following: can literature expand our imagination?

As she convincingly shows at the very beginning of her argumentation, the idea that literature can expand or sharpen our imagination in ways that make readers appreciate inedited experiential states is widespread among writers. Just to mention a case among many others, T.S. Eliot wrote:

> [The poet is] not merely a more conscious person than the others; he is also individually different from other people, and from other poets too, and can make his readers share consciously in new feelings which they had not experienced before. (Eliot 1957: 9)

It is interesting to notice how this idea has enjoyed some success not only among novelists and poets, but also among philosophers. Here as well I mention only one case that might be taken as exemplary of this position. Martha Nussbaum wrote:

> [Literature educates] not only horizontally, bringing the reader into contact with events or locations or persons or problems he or she has not otherwise met, but also, so to speak, vertically, giving the reader experience that is deeper, sharper, and more precise than much of what takes place in life. (Nussbaum 1992: 48)

Peacocke's argumentative goal is in line with these writers' and philosophers' considerations. Some of the arguments that she presents to substantiate the claim that literature can expand our imagination can be extrapolated from the context of literature and applied to empathy in general. As we will see,

even if we agree that literature constitutes a privileged *locus* in which imaginative expansions can take place, empathic engagement with other people's experiences could also succeed in achieving this goal.

Some of Peacocke's arguments resonate with concepts and ideas we are already familiar with. She first individuates a specific kind of imagination, namely *phenomenal imagination*, which is a kind of imagination that is distinct from other kinds of imaginative activities because it has phenomenal properties as its main content. Roughly put, phenomenal properties consist of feelings, i.e. ways in which experiences *feel like* "from the inside" (Peacocke 2020: 4). A phenomenal property is captured by a *phenomenal concept*, which allows us to both i) recognize the relevant phenomenal property when it occurs in experience and ii) imagine what it feels like to have an experience with that phenomenal property.[63]

In light of this, we can conclude once again that in order to imagine what an experience feels like, we need to have undergone an experience that instantiated the relevant phenomenal property we want to recall. Peacocke argues that phenomenal concepts guide phenomenal imagination, but that sometimes phenomenal imagination, when prompted in the right manner, can make us acquire new phenomenal concepts.

5.2. Noticing

In order for us to build a phenomenal concept of a phenomenal property, we need to *notice* that phenomenal property. The point is fairly intuitive and generally uncontroversial: it is simply not sufficient for us to have an experience of a certain phenomenal property to build a phenomenal concept of it.

> To **notice** some phenomenal property is (inter alia) to attend consciously to an instance of that property. To notice a phenomenal property is thus to establish a *de re* cognitive connection with an instance of it that supports demonstrative reference to the property in question—for at least as long as your experience has that property. (Peacocke 2020: 7)

This implies that we certainly do not notice every phenomenal property we are confronted with in our experience. It is precisely in virtue of this that our imaginative representational capacity can be

[63] David Lewis would have added that the information we get through experience (what we called with Peacock a "phenomenal concept") allows us to also *recall* what it felt like to undergo that experience (D. K. Lewis 1990).

expanded when rightly prompted. Literature, so Peacocke argues, is precisely the tool that guides our attention in a way that makes us notice new phenomenal properties we did not notice before, endowing us with new imaginative capacities. By making us aware of unnoticed aspects of our experience, literature provides our imagination with new phenomenal concepts, hence expanding its representational scope.

As I stated in the previous section when talking about the notion of experiential competence, noticing *simpliciter* is not sufficient for us to expand the range of our phenomenal imagination: to do that we need to *store* that information in our memory, so that we can establish a longer-term capacity to represent the relevant phenomenal property even in the absence of a *de re* connection with it. In other words, while merely noticing a phenomenal property might make us aware of it *at the time we are having the experience*, a phenomenal concept stores that information so that it can be retrieved and deployed even in the absence of the relevant experience.

At this point, one might argue: how is it possible for us to acquire new phenomenal concepts via phenomenal imagination if phenomenal imagination is by definition enabled by phenomenal concepts? In fact, to imagine what it feels like to taste coconut, we need to have the phenomenal concept of the taste of coconut, otherwise our imaginative project would simply fail to take off.

To this we can reply by referring to what I have already said about imagining under constraints. We can either *actively* produce imaginings of states or events that we already know – like when we imagine tasting a coconut – or be *passively* guided by external sources of information that constrain our imaginative representations in unplanned ways. Take the case of reading a novel. In this case, what we imagine is guided by what we are reading. In this case your imagination is *constrained* by information coming from an external source. This partly frees the imagination from its dependence on a set of pre-existing phenomenal concepts so that it can be led to unexplored representational territories.

While reading certain stories or particularly apt descriptions of experiential states, literature is capable of redirecting our attention to past experiences and make salient for the first time previously unnoticed phenomenal properties.

5.3. Comparisons and involuntary memory

Peacocke claims that involuntary memory is what allows our imagination to aim at previously unnoticed phenomenal properties of our past experiences. She traces back the notion of involuntary

memory to the work of Marcel Proust. In a nutshell, involuntary memory is a memory of our past experience that is triggered by an external cue and over which we have no control. It is capable of evoking a remarkably rich and detailed phenomenology that can in principle overflow the boundaries of the repertoire of our phenomenal concepts (Peacocke 2020: 12).

We do not have to delve into the details of Peacocke's interpretation of Proust's concept of involuntary memory in order to appreciate its implications. Literature can trigger memories of our past experiences (what Proust called involuntary memory) and then make us realize, by means of comparison, previously unnoticed phenomenal features. Of course, this cannot happen when only one involuntary memory is brought to mind, but only when two or more of them are triggered by the literary text.

Comparing two (or more) distinct sensations linked to different involuntary memories prompts us to search for a *common ground*, i.e. the phenomenal aspect that these experiences share. As a result, we can get a new phenomenal concept. Literature is rich of relevant examples. One of the most effective reported by Peacocke is the following one:

> Every street lamp that I pass
> Beats like a fatalistic drum… (T.S. Eliot 1920)

By comparing the light of the lamps on the street to the sound of a drum, Eliot draws our attention to a feature of that particular visual experience that we are unlikely to have noticed before. Every light feels like hits of a drum that punctuate the rhythm of our steps. This passage evokes representations two kinds of experiences, i.e. walking by the streetlamps and hearing the sound of drums and prompts us to search for a common ground between these two experiential states.

Or take one of my personal favorites:

> Yet, one day through an open gate,
> among the green luxuriance of a yard,
> the yellow lemons fire
> and the heart melts,
> and golden songs pour
> into the breast
> from the raised trumpets of the sun (Montale 1925/2002)

Here, the poet is making us aware of a specific quality of the sunlight, which is compared to the sound of trumpets. As in the previous case, we have a comparison between two sense modalities which

points at a previously unnoticed aspect of an experience. The visual experience of sunlight is compared to the high-pitched, merry sound of trumpets. Again, that endows us with a new phenomenal concept, one that allows us to notice a feature of sunlight of which we were not aware before.

These literary devices constitute a way of framing a certain aspect of experience within the gradient of a novel perspective, that can make salient previously unnoticed phenomenal properties. These properties can be stored in new phenomenal concepts that can then be deployed in future experiences as a recognition-device – e.g. we could be reminded of the sound of trumpets whenever we will see a particularly sunny day – or in further imaginative projects – e.g. we might imagine a specific character of sunshine, namely one that can be compared to the sound of trumpets.

5.4. From literature to empathy

The previous literary examples merely served the function of highlighting the way in which we can expand the scope of our imaginative by letting our imagination be guided by external triggers and sources of information. In the case of literature, it is the literary text that prompts our imagination in an epistemically generative manner.

However, I firmly believe that we can extend some of the arguments presented above outside the scope of literature. There are other, non-literary cases in which our imagination is prompted to represent experiential features that we were not previously aware of.

More specifically, I believe that being confronted with the non-fictional, real-life stories of other people can prompt our imagination in the required manner to expand its epistemic scope even without the sophisticated tools of comparisons, metaphors, and similitudes deployed by literature. In other words: people can make us aware of previously unnoticed aspects of experiences.

Other people's experiences can trigger our imagination so that it can achieve representational goals that were not available to us before we actually engaged with them. Coming back to our previous example, when our friend told us that her fight with her father generated a surprising state of relief in her, we might realize that, in some circumstances, ending relationships with close ones may result in feelings that are different from those we were reasonably expecting. Instead of (or alongside) despair, anger, and frustration, it is possible to experience relief. A complicated relationship can result into feelings of liberation and reduce the amount of stress that comes from them. Moral: our friend's story

can teach us that break-ups – of any kind – might be processed through a (for us) unexpected emotional lens, i.e. it teaches us a new way in which break-ups can be experienced.

But now we can ask: how can we make such a discovery? Peacocke's suggestions are again useful in this context as well. Here as well, indeed, a kind of comparison is taking place. It is by means of comparison that I can grasp the relevant phenomenal quality of my friend's experience in the previous example. By imaginatively juxtaposing the anger towards a loved one and the relief for the end of something that was emotionally poisonous for us, we can get closer to grasp the phenomenal character of her experience. Memories of our past experiences of anger and relief are triggered by her testimony and our imagination is therefore prompted to represent something unexpected: we are representing a state of affairs that were not able to grasp before actually engaging with our friend's story.

6. More on empathy and memory

If we take a closer look at the arguments presented above, Peacocke is simply claiming that imagination can be put to work so that it can target unnoticed features of experiences we already had. She highlights the important role of (involuntary) memory in bringing about phenomenal details of past experiences we had that can overflow the set of phenomenal concepts at our disposal.

However, the kind of process she is describing remains slightly ambiguous. On the one hand, we have the phenomenally rich recollection, through involuntary memory, of past experiences that can be so vivid to overflow the repertoire of phenomenal concepts at our disposal, and on the other hand, a process of imaginative expansion that is achieved by means of comparison, i.e. by juxtaposing representations of two experiences to find a common ground that can provide us with a new phenomenal concept.

I believe that these two processes can come apart and are mediated by two distinct kinds of episodic recollections that can either work in tandem – as Peacocke's account seem to assume – or in substantial independence from one another. These two recollection processes are:

(i) *Re-inspection of past memories*.

This process consists in the recollection of a past experience we had and that can be re-inspected in a way that can make us aware of a previously unnoticed phenomenal feature. Let us come back to the case of our friend and her father. Her testimony might trigger the recollection of a similar experience we had in the past and make salient to us features of it that we did not notice before listening to our friend's story. I might be reminded of that

time in which I also suddenly ended a relationship with a close one and, as a result, felt some sense of relief. It is perfectly plausible that, back at that time, I was not aware of such a relief that came with the break-up and that only once I was suitably triggered by my friend's story, I came to realize how that experience felt.

Notice that holding this view does not commit us to the existence of unconscious emotions. I am not claiming that we were unconsciously experiencing relief, but merely that we did not *notice* that we were. Before engaging with our friend's story, we might well have framed our break-up purely in terms of anger, sadness, and frustration. But after re-inspecting that experience, in light of our friend's story, we might be able to spot for the first time a different affective shade of it.

(ii) *Recasting of past memory traces*.

In this case, rather than recollecting a past experience and re-inspecting it in light of a new interpretative framework, we are recombining features of different past experiences, so that we can come with representations of experiences we never had before. Differently from the previous case, in this one we are simply recasting memory traces of past experience for a new imaginative project that might eventually lead us to appreciate the character of an experience which we did not undergo as such.

Peacocke conflates these two processes and therefore fails to appreciate the specificity of both. While she is willing to claim that we can combine memories of two or more past experiences (ii), she also maintains that the imaginative expansion is something that relies on the recollection of experiences we already had and that is therefore a mere consequence of a re-inspection of them under a new "salience filter" (i). These two processes can certainly work in tandem: I can re-inspect two or more memories and then compare them in a way that provides me with a new phenomenal concept. In Eliot's passage I mentioned in the previous section, we recollect two experiences we already had, i.e. the experience of walking down a street with streetlamps and the experience of the sound of drums, and we compare them so that we build a new phenomenal concept, i.e. the drum-beat-like character of streetlamps.

However, process (i) and process (ii) can also come apart. And this is precisely what allows us to make an even stronger claim than the one made by Peacocke, namely that imaginative expansion is not limited to the mere re-interpretation of experiences we already had. When properly triggered, imagination – and, therefore, empathy – can achieve a representation of experiences we never had in the past.

If we look closer at the process described in (ii), we can actually see how it allows our imagination not to be bound to the mere re-inspection of similar experiences but can combine traces of old ones in order to represent new ones.

Let us come back to the chickenpox example: does the empathizer really need to have contracted the same disease or can she work out what it feels like to have chickenpox without having contracted it herself? Indeed, she might surely have experienced some occurrences of itchiness and some others of frustration even though they maybe did not happen both at the same time and for the same reason as in the case of chickenpox. In this latter case, it seems that there would still be space for the empathizer to understand what it feels like to have such a disease even though she did not experience it as such.

Of course, that would be cognitively costlier because instead of recollecting the experience as such from our memory, we would have to somehow "re-assemble" it in our imagination, so to speak, from scratch. Nevertheless, this could in some cases amount to a mere pragmatic difficulty, rather than constitute a principled impossibility.

The above considerations are far from being purely speculative but are in harmony with some recent findings in the neuroscience of memory. In fact, our capacity for episodic memory seems to be subserved by a larger system in the brain that is also recruited for a variety of purposes such as imagination, dreaming and social cognition (Hassabis and Maguire 2009; Mahr and Csibra 2018; Schacter et al. 2007; Spreng 2013). The contemporary literature in philosophy of mind and cognitive science is immensely enriched by a lively and intriguing debate about the nature of episodic memory, e.g. what distinguishes it from other forms of remembering, how we should conceive of the relationship between memory—more specifically, episodic memory—and imagination, just to mention a few issues.

Without committing ourselves to any strong account of episodic memory, we can simply draw a lesson from this debate: information stored in our memory (i.e. memory traces) can be flexibly recast for a variety of purposes, one of which— not the exclusive one —is episodic memory. That implies that the same contents that figure in episodic memory can be recruited and flexibly recombined in imagination, and, more specifically, in empathic imagination. Being the result of past experiences that we personally underwent, memory traces can vehicle information about the experiential features of a specific event, i.e. they can convey information about what it feels like to undergo a certain kind of experience. Memory traces of our past experiences can therefore be recruited in order to serve as "raw material" on the basis of which we can calibrate our imaginative simulation of other people's perspectives. Memory traces do not amount to the actual recollection of a fully-fledged episodic

memory but are rather the building-blocks of representational processes subserving both episodic memory and imaginative (or empathic) representations.

Again, this idea is largely in line with a research thread in cognitive science that regards episodic memory and various imaginative processes (e.g. perspective-taking and empathy or counterfactual thinking) as two different outcomes of the same general cognitive process. This idea has been recently dubbed as the "constructive episodic simulation hypothesis" (Mahr 2020):

> According to the 'constructive episodic simulation hypothesis' […] episodic simulation functions by retrieving and *flexibly recombining* elements of stored perceptual information (i.e. memory traces). (Mahr 2020: 3, emphasis added)[64]

Addis states the core principle of this hypothesis even more clearly:

> Remembered and imagined events are the same type of representation in that both are simulations of experience constructed from the same pool of experiential details and under the guidance of the same schemas. While there may be some differences in the relative proportions of content types, with remembered events typically containing more perceptual content than imagined events for instance [...], this is only a matter of degree because the types of details comprising these representations are overwhelmingly the same. (Addis 2018:3)

The constructive episodic simulation hypothesis is certainly fascinating and largely in line with the arguments I am presenting in the present section. However, we do not have to bring on board the entire set of assumptions and implications of this theory in order to bring home the lesson I want to draw. The only thing we need is to accept the intuitively persuasive and empirically solid idea that memory traces of past experiences can be recruited and flexibly recombined for novel imaginative projects, e.g. empathic perspective-taking.

What I have said so far, though, should by no means be taken as if empathic understanding amounted to *nothing but* the mere recombination of our previous experiences projected onto our target. What I am defending instead is the much plainer and more intuitive claim that imagining other people's experiences is enriched and partly constrained by our own past experiences and, more precisely, by what I called experiential competence. Indeed, the claim that our empathic capabilities are

---

[64] This quote seems to restrict the scope of memory traces to mere perceptual information. However, we do not have to buy such a restrictive view: we can talk about other kinds of information stored as a memory trace, e.g. affective or emotional character of a certain event. I do not have space to further argue about this aspect. In any case, this issue is not central to the point I am trying to make here.

constrained by what we have personally experienced plus our ability to retrieve it, and the claim that our empathic understanding *reduces to* what we have experienced in the past, should carefully be distinguished and kept aside in order to avoid conceptual confusion. The flexibility with which stored information of our past experiences can be manipulated and recombined allows us to produce empathic outputs that could not be plainly reduced to the experiential competence we possessed before engaging in the empathic process. That is what ultimately grounds what we called empathy's epistemic generativity, viz. empathy's capability to provide new knowledge about experiences we did not personally undergo. Said more clearly: modelling our empathic imaginings alongside our target's experiences is in fact what allows our representational capability to *stretch* and recombine our experiential competence so that we can achieve some new knowledge. Acquaintance with our targets' experiences can enable us to represent experiential features for the first time, in an inedited fashion.

A concrete example might help us illustrate this point more clearly. Suppose that I never felt a joy so intense to drive me to tears. That implies that I lack the specific experience of crying out of joy. Still, I experienced joy and I cried sometimes. My acquaintance with cases of people crying out of joy provides me with some evidence that allows me to notice in a new fashion a personally unexperienced feature of joy. This external acquaintance enables me to understand how it is possible to experience a joy that is so intense to lead to tears and what that experience might feel like: I can flexibly recruit my experiential competence in order to produce an empathic output that enables me to combine together the powerful impact of joy and the personal unsettlement of the experience of crying. Of course, as already stated, the fact that we can flexibly manipulate our experiential competence does not mean that it is completely unbound. The complete lack of the relevant type of experience would imply a total empathic blindness

Still, most of the time, most adults are provided with a broad experiential palette that allows us to expand the scope of experiences we can empathize with. As we saw, flexible reuse of our memory traces and acquaintance with the specificity of our target's perspective can in fact enable us to push our empathic capacities beyond what we already know about ourselves.

7. Personal Outcome: Changing through empathy

So far in these last sections, I have focused on the strictly epistemic consequences of empathy. I tried to show that empathy, by letting us engage with different perspective can *teach* us something new

about certain experiences or phenomenal features we either did not experience before or were not fully aware of them before actually embarking in the empathic project.

However, empathic perspective-taking is not limited to the mere increase in the body of knowledge at our disposal. Or better, the newly acquired body of knowledge can actually affect us in ways that go beyond a mere epistemic growth. Indeed, by coming to appreciate certain facts about other perspectives which we ignored before engaging in the empathic process, we might undergo changes in our behaviors and attitudes. Most interestingly, these changes can be enduring and last long enough to change the way we are and in which we perceive ourselves in relation to others and to the world around us. Stated otherwise, they can be personally transformative.[65]

In the following sections, I am going to explain the ways in which empathic information, i.e. the kind of knowledge and understanding that we are capable of acquiring by engaging in empathic processes, can reach beyond the boundaries of epistemology and affect us as individuals.

7.1. The empathic framing of information

Throughout the present work it might seem that I have based many of my arguments on an implicit assumption, which I never fully disambiguated but that clearly was grounding the efficacy of some of the most ambitious theses that I defended in the previous chapters. The supposed implicit assumption is the following one: the understanding of other minds that is achieved through empathy is *distinctive* and, in principle, *unavailable* via other non-empathic representational tools. If I want to represent what it might feel like for my friend to dive into the ocean, I have imagine myself doing the same thing and try to model the phenomenal details of that experience – the cold water on my body, the shortness of breath and at the same time the euphoric excitement that results from it, etc. –, in other words, I have to *empathize* with my friend's experience.

It seems that knowledge about distinctive phenomenal features of lived experiences can be rehearsed only if we embark on an empathic process. I personally do not have anything against this view and I very much sympathize with it. However, others might be less enthusiastic about the idea of committing to such an extreme hypothesis. One might not be so peacefully inclined to believe that there are certain experiential facts that cannot be apprehended from a, so to speak, third-personal

---

[65] This does not imply that every personal transformation is grounded in an epistemic transformation. As we saw, the two aspects are strictly intertwined, but they can definitely come apart.

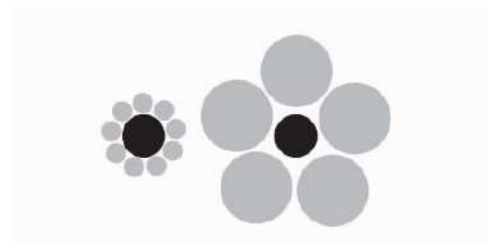
perspective. Even if one wants to remain open about the possibility of accessing certain fundamental features of experience that I took to be essentially linked to empathic modes of access, one can still appreciate most of the arguments I presented above and those that I am about to defend. Indeed, even if one maintains that empathic information could be, in principle, achieved by other non-empathic, third-personal modes of presentation, one should still be willing to acknowledge that empathy might be the likeliest and most natural process through which we can gain such information.

With respect to this particular issue, Steinberg (2014) has argued that empathy represents a *user-friendly* way to convey relevant information about certain distinctive features of lived experience. Therefore, we are not forced to buy any argument in support of empathy as an exclusive and irreplaceable form of understanding in order to appreciate its epistemic and personal implications:

> I think that there are good reasons to believe that experience [and, therefore, empathy as a quasi-experiential mode of apprehension] is not the exclusive reservoir of certain kinds of facts, but rather is a particularly colorful and user-friendly mode of apprehension. (Steinberg 2014:7)

By user-friendly, Steinberg means that the empathic framing of information can promote the satisfaction of one's goals by nudging our practical reason in ways that are more functional to the fulfillment of such goals.

In order to illustrate this point, psychologists usually resort to some examples of visual illusions like the Titchener or Ebbinghaus illusion:



Most of us perceive the black circle on the left as being bigger than the one on the right. However, the two circles are completely identical. The mere fact that we draw the same circle surrounded by smaller circles makes it appear bigger than it would if we were to draw it as surrounded by bigger circles.

Now, this example is of great interest for philosophers and psychologists interested in perception, but it can also teach us a general lesson about how the mind works. The way in which the same information is presented, remarkably changes the way in which we apprehend it. In the above case

presenting the dark circle within a context of smaller circles makes it appear bigger than its identical counterpart that has been presented within a context of larger circles.

This phenomenon has widely researched consequences for our decision-making. For instance, relevant works in behavioral economics have shown that the way in which we frame the same information, can lead to remarkably different effects on the subjects' decision-making. For instance, a classic experiment (McNeil et al. 1982) showed how people are much more willing to undergo a certain medical treatment if they were told that they had 90% chances of surviving, but much less willing to undergo the same treatment if they were told that they had 10% chances of dying. In this case, again, *exactly* the same information is conveyed but people responded differently to different ways in which it was presented.

The same line of reasoning can be extended to empathy. Information that is accessed through empathy tends to be processed differently from other, non-empathetic modes of appraisal. The experientially vivid representations of empathy frame information concerning other people's minds in ways that make certain features more salient to the empathizer. This, in turn, can have remarkable repercussions on empathizers' practical reasoning and subsequent behavior. Conversely, certain features of other experiences would likely be silenced or cut-off altogether via other, non-empathic means of appraisal.

Let us take a practical example in order to illustrate this last point. Suppose we are trying to stop eating chips because we decided to go on a diet. One possible way of dealing with the temptation of eating chips consists in avoiding any cue that might trigger vivid imaginings of us eating them. We know that eating chips is an intrinsically pleasurable experience for us, but we are committed to this diet and as long as we keep a disengaged stance towards eating chips, we can manage to avoid it. However, as soon as we engage the vivid imagining of eating them (we imagine their salty taste, their crispiness, etc.), we feel the temptation of eating them ramping up. A lot of features of the experience of eating chips that were "silenced" by our disengaged approach towards them suddenly emerged: a wholly different take on the same experience is instantiated.

We have at least two modes of appraising the experience of eating chips: one in which we focus on our long-term goal and maintain that the eating chips, however pleasurable, should be avoided, and another in which we rehearse through imagination some of relevant features of the experience of eating chips. I assumed that in the second case we would have a much harder time resisting the temptation. But this is not the moral of the story, the actual lesson that I want to draw from this example is the different impact that two ways of appraising the same thing – i.e. eating chips – can have a different impact on our practical reasoning and subsequent behavior. This is because different

features of the same thing are put under a different of salience filter: in the first case, one that emphasizes the bad impact that eating chips can have on our health, in the other case, one that channels our attention towards the very experience of eating chips.

Justin Steinberg has another very persuasive example:

> Consider an ordinary case of *akratic* action. Imagine that I have a weakness for Amaretto cookies, but I also have a moderate, though not life-threatening, allergy to almonds that are in them, I might sincerely judge that it is in my best long-term interest to avoid them, even while giddily devouring a handful of them. One way of explaining what is happening in this and many other self-regarding instances of *akrasia* is that I am hyperbolically discounting the future. And this hyperbolic discounting may itself be understood as a failure to adequately take up the affective perspective of my future self – I have not sufficiently empathized with *him*, or taken up *his* perspective. I might have thought to myself, 'well, I'm going to wind up regretting this,' but the reasons for regret were apprehended in a way that is too muted to focus my attention. […] Had I adequately empathized with my future self so that I could *now* feel what it will be like to break out in hives and so forth, this information might have been sufficiently vivid to offset the present temptations – I might have been able to fight fire with fire, so to speak. But without empathy for my future self, the affective force of the future consequences is too weak. (Steinberg 2014: 56)

The main issue with these cases can be convincingly attributed to the way in which a certain kind of information is apprehended: empathetic vs non-empathetic modes of appraisal. The fact that I know *that* certain actions are going to be bad for my future self does not guarantee that I might act in ways that allow me to avoid such future threats. This may be due to the fact that, by failing to fully empathize with the perspective of my future self, I am not sufficiently pushed to act in ways that are congenial to his best interests. By this, I am not claiming that full empathy with one's future self would *necessarily* result in a behavior that maximizes its best interests. What I am trying to show is that empathy does indeed have a distinctive impact on our decision-making and behaviors, for better (like in the amaretto example) or worse (like in the chips example).

It should be obvious that the above considerations do not merely apply to the *intra*-personal case, in which we are asked to evaluate two different perspectives at two different times, which are

nevertheless occupied by the same subject, viz. us. These arguments extend to the *inter*-personal case, in which it is also true that we can *discount* perspectives which we do not access empathically.[66]

These considerations simply aim to show how empathy can impact subjects in a way that can change how they process information and subsequently behave in light of it. Now, the next step consists in seeing more concretely how the impact that empathic information can exert on the empathizer can bring about an actual personal transformation.

7.2. Brief clarificatory interlude

A brief disambiguation is now in order. From what I said so far, it might seem, indeed, that I am surprisingly avoiding mentioning the moral implications of the considerations above. This is a deliberate choice. Indeed, as stated at the beginning of the present work, my focus here does not directly concern the moral implications of empathy, but mainly targets its epistemology. What I have been concerned about throughout the present work is what kind of knowledge and what kind of understanding is empathy capable to achieve. This, I believe, is a first crucial step before we can engage in any sound and productive discussion about empathy's supposed moral powers.

However, engaging very briefly with some moral concerns that seem to spontaneously emerge from the previous section, will allow me to better qualify some of the considerations presented above and introduce the topic of the next section.

One might rightly notice how the above example about eating chips highlights one of empathy's main moral flows. Indeed, by representing the experience of eating chips from a first-person perspective – i.e. under what we might call the empathic mode of presentation – some of the relevant features of that experience became salient. Unfortunately, those features that are made salient are precisely those that compromise the fulfillment of our best interests, i.e. in this case, avoiding junk food. By vividly imagining the experience of eating chips, I am much more prone to fall prey of the temptation to eat them.

---

[66] This is precisely the direction indicated by renowned experiments carried by Batson and its team, in which they managed to consistently show how people who are prompted to empathically take up the perspective of subjects in need of help, do show a much higher tendency to help (see *Chapter I*).

It is easy to see how this might easily translate into a stringent moral issue. Psychologist Paul Bloom, probably better than anyone else, has evidenced this flaw, which he calls empathy's spotlight nature (Bloom 2016).

> "Empathy is a spotlight focusing on certain people in the here and now. This makes us care more about them, but it leaves us insensitive to the long-term consequences of our acts and blind as well to the suffering of those we do not or cannot empathize with" (Bloom 2016: 47)

Empathy channels our cognitive resources and moral concerns towards those with whom we are empathizing. This results in a disproportionate allocation of care towards the targets of empathy and can drive us away from some greater goods that escape empathy's narrow scope.

Similar concerns have also been voiced by Jesse Prinz who complains about empathy's narrow scope of interest, which he takes to offer a secure ground for further inequalities and moral wrongdoings:

> When we empathize with a person awaiting an organ transplant, we let her jump to the front of the queue, elbowing out many who have been waiting longer. Likewise, an empathetic plea for hunger relief might cause us to send checks to one family rather than a village, or we might help one community, when others are in greater need. (Prinz 2011: 228)

Bloom and Prinz are certainly dealing with a broader notion of empathy that does not perfectly coincide with the one I deployed in the present work. However, their critique can be easily extended to the present account of empathy as well. Here as well, the salience with which we represent the suffering of the person we empathize with might *blind* us, channeling our cognitive resources and concerns exclusively towards the person we are empathizing with. We fail to balance empathy out with a more cold-blooded approach to the matter and this might result in short-sighted judgments and behaviors.

This is certainly a troublesome feature of empathy and I believe no one should deny this. The disproportionate emphasis empathy places on other's experience can actually push us to behave in ways that would collide with our best (in this case, moral) interests.

However, there are at least three considerations one should keep in mind when approaching this kind of critique. They certainly do not fully dismiss the power of the spotlight critique, but certainly give us some reasons not to consider it a knock-down argument against empathy. At the same time, they allow me to re-state and better contextualize some of empathy's core epistemic characteristics.

(i) *Empathy is an antidote against discounting.*

In some cases, salience is *precisely* what we need in order to be motivated to act. As we saw in the previous case involving the Amaretto, being able to accurately empathize with the perspective of our future self, is what is likely to inhibit an inconvenient temptation for Amaretto: by making salient the negative impact that eating the Amaretto could have on our future experience, we might have one powerful reason more not to eat it. Empathy may brilliantly work as an antidote against *discounting* perspectives: by making us aware of what it feels like to inhabit them, it provides us with relevant knowledge that we might be likely to leave out when we approach them through other non-empathic modes of appraisal. Again, I might simply know that Amaretto is bad for my health, however in my evaluation I might leave out or discount *how bad* it can be for my health, i.e. all the painful experiential states I might undergo as a consequence of eating them.

(ii) *Empathy can disclose new "types" of perspectives.*

Empathy might brilliantly work as a "starting point" for disclosing issues we were not sensitive about, i.e. it might be the best way to make us aware for the first time of e.g. a certain moral wrongdoing or injustice. By doing this, empathy makes us sensitive to a perspective that is not restricted to the individual case but can be taken as constituting a generalizable way of looking at the world. By empathizing with, say, a member of an oppressed minority we might be endowed with the relevant knowledge that can make us sensitive to these issues. We are not forced to obtusely stick to the single person but can take empathy as an open door towards a much larger issue. The individual we empathized with can be taken as representing a certain *type* of individual in a certain *type* of situation.

(iii) *Empathy is situated in a larger cognitive context.*

Finally, empathy is not something that we do in *a vacuo*, i.e. it is not an all-or-nothing process that exerts supreme cognitive monopoly on us. Empathy is not something that happens in complete isolation from other processes and reasoning routines. It can be balanced with other strategies we might deploy – e.g. utilitarian cost-benefit reasoning, etc. Many critiques of empathy seem to rely on the implicit assumption that once we empathize, there is no way back, we cannot counterbalance or integrate empathy's outcome with other processes. But in many circumstances, multiple approaches to the same matter is precisely the best strategy we have.

A more in depth-discussion of these issues would bring me to far afield. As stated above, these considerations merely served the purposes of addressing some urgent concerns one might have about empathy and morality and introducing the topic of the next – and final – section.

7.3. Empathy and other ways in which the world can be

I want to highlight at least two ways in which we can say that empathy can change us personally, namely occurrently or dispositionally. Empathy can change us occurrently in the sense that it can affect the way in which we think, feel, and act in certain circumstances. Examples are easy to bring to mind and largely studied in experiments that I already cited multiple times in the present work. For instance, if I am prompted to empathize with someone in need of help, I might come to feel a certain amount of distress and decide to help them. From this, it is safe to conclude that empathy *occurrently* caused a change in my subsequent feeling, thinking and behavioral patterns. However, empathy can also have some more enduring consequences on us that go beyond its transient influence on certain occurrent mental states or behaviors of ours. Empathy might change the way we might are *disposed* to think, feel, or behave in the future. Stated otherwise, it might change us *personally*, (partly) revising the kind of people we are.

It is precisely this latter sense that interests me here. The fact that empathy can change us dispositionally, thus transforming the kind of individuals we are, can be directly drawn from the characteristics of empathy that I have been explaining throughout the present work.

However, in order to set the issue on the right path, it might be useful to introduce one last useful distinction. In a recent paper on the relationship between empathy and testimonial injustice, Olivia Bailey (Bailey 2018) individuates two different senses in which we can conceive of other people's testimonies about certain experiential states they are in.

(i) A *weak sense*, in which we take people's testimony simply as a certification of a certain psychological state they are in.
(ii) A *strong sense,* in which we treat other people's testimony as accounting for certain features of the world outside them.

As usual, a concrete example might help. Let us imagine that our friend Andrew is afraid of dogs. Whenever he is confronted with a dog, he shivers in fear. Imagine that Andrew voices out his state and shouts out: "that dog is terrifying". According to the above distinction, there are at least two ways in which we can take Andrew's testimony. If we take it in the weak sense, we might simply interpret

it as stating a certain state of fear and discomfort Andrew is in. He is afraid of dogs and undergoes negatively valenced emotional experiences whenever he is faced with one. If we take it in the strong sense, we interpret it as reporting an actual state of affairs in the world, i.e. one in which the dog actually looks terrifying. When I claim that in this stronger sense, we treat Andrew's testimony as certifying a certain state of affairs in the outside world, I do not mean that we take it as referring to some mind-independent property out there. Of course, the dog is not *terrifying* in itself. However, if we succeed in looking at the dog through Andrew's perspective, it is certainly *the dog* that is terrifying. Andrew's state tracks on certain evaluative facts that are perceived as belonging to the world out there – in our case, to the dog.

Of course, this logic is not limited to the mere case of testimony but can be extended to the way in which we treat others' perspectives in general. We can take someone's perspective in the weak sense, as the simple conglomeration of certain mental states that are instantiated by the subject occupying it. Or we can take someone's perspective in the strong sense, i.e. as a way of looking at the world, which can pick up on certain relevant features of things out there. To take someone's perspective in the weak sense, full intelligibility of the other perspective is not required – i.e. we do not have to be able to see the dog as actually terrifying. To take someone's perspective in the strong sense, instead, we need empathy, i.e. we need to be able to see the dog as terrifying, as *warranting* a reaction of fear.

In light of this, it seems possible that, as a consequence of an act of empathy, we might revise our previous preferences in ways that might drive us away from the previous self we were before. If we manage to see the world through the lens of an alien perspective, we might come to appreciate a new way in which the world may be. As a result, we might, so to speak, be "swallowed" by the newly appraised perspective.

My considerations about empathizing with a desire turn out to be particularly useful at this point. If being able to first-personally imagine holding a desire requires that we imaginatively bring us to see what is *desirable* about a certain goal we might have, then we expose ourselves to the possibility of actually coming to hold such a desire. Via empathy, we might become aware of certain desirable features of the object of desire and find out, at the end of the process, that our preferences have changed.

If we want to understand why certain people might hold the desire to climb Mount Everest, we might start to enquire about their reasons for having such a goal. Rather than *discounting* these people's desire as a somewhat quixotic tendency of bizarre individuals, we might want to figure out the what it might feel like to hold such a desire. We might then decide to listen to their testimonies, e.g. stories

about how they developed the ambition to climb the highest mountain in the world. They might tell us about the excitement for such an extreme challenge, some others might describe the feeling of fulfillment and awe they experienced once they reached the top. By trying to capture these people perspectives, we might start to look at things through their eyes, and we might start understanding what might be so desirable about such a challenge. As I said in the previous section, empathy can disclose for us a new *type* of perspective, one that we might actually endorse from that moment on. This process could then result in us actually developing a passion for climbing: we might start taking climbing lessons and going on hikes over the weekend.

This example shows us how we might be changed not only occurrently, but also dispositionally by an act of empathy. By making an alien perspective intelligible to us, in the strong sense defined above, not only we avoid the issue of discounting it, but we actually expose ourselves to the risk that it might become our own. It might disclose a new way of looking at the world which can actually restructure our egocentric map, e.g. by endowing you with a new set of preferences, values, beliefs, or desires.

A final disambiguation is in order. Accepting the idea that empathy *can* personally transform us, I am not claiming that it *must*. The fact that I can come to empathically access a desire or a preference, does not entail that I necessarily come to hold it. This is in virtue of the third feature that I highlighted in the previous section. Besides making a certain perspective salient and besides being capable of disclosing a new way in which we might look at things, empathy is still *situated* within the larger context of our psychology. This means that empathic processes can be counterbalanced by other processes or kinds of reasoning we might engage in. For example, even if I come to see what might be so desirable about climbing Mount Everest, I might still decide not to embark in such a life-threatening challenge because e.g. I might simply be too afraid of such an extreme challenge or I might have other preferences of higher value for me. I might, for instance, have a solid pre-existing desire to become a football player and therefore come to the conclusion that a career as a climber is simply not available.

In any case, this does not change the general moral of the story. Empathy *can* – even if it does not have to – can be a personally transformative experience. By disclosing alien perspectives to us, it also exposes us to the possibility of reshaping the kind of self we are.

Conclusions

End of the Journey (Looking Ahead)

> In all people I see myself, none more and not one a barleycorn
>
> less,
>
> And the good or bad I say of myself I say of them ("Song of Myself", Walt Whitman)

Seeing oneself in "all people", as Whitman does in this poem, can be interpreted in at least two ways.[67]

On the one hand, it could be seen as the ultimate fulfillment of empathic ideals, as the accomplished "community of souls" I mentioned in the *Introduction* to this work. Other people cease to be strangers and become as familiar to us as we are to ourselves. The empathy gap is finally bridged.

On the other hand, it could be interpreted in a more pessimistic way. It can be seen as the ultimate triumph of psychological egocentrism. No matter how hard we try, we will never be able to imagine our way into the perspective of someone else, we would always be bound to ourselves, to the kinds of people we are, with no hope to bridge the empathy gap.

This brings us back to where we started, between Coetzee and Proust. The present work is an attempt to show how empathy sits precisely between these two extremes, pulled from one side to the other in its attempt to provide us with an insight into other people's minds. From this tension spark both empathy's epistemic limits and possibilities which I (partly) surveyed in the previous chapters.

The main take-home message is that empathy's epistemology cannot be adequately captured by means of an "all-or-nothing" logic. Deeming empathy as hopelessly defective would be as inaccurate as cheerfully acclaiming its unbound powers of understanding. Only a careful and nuanced weighting of what empathy can and cannot do can give us a truthful picture of what it means to imagine "our way into another being", to take up alternative perspectives from the one we usually inhabit.

The present work, in its attempt to explore empathy's conceptual and psychological profile, will have met its biggest ambition if it succeeded in showing how the constant struggle between understanding

---

[67] Here I do not aim to be faithful to what Whitman actually meant, I simply take his poem to be open to *at least* the two interpretations I offer.

and misunderstanding, between self and other, egocentrically biased and accurate perspective-taking, actually constitutes the very essence of empathy.

Before concluding this work, it might be useful to spend just a few lines looking ahead at the future research that could be done in light of the epistemic gains of this thesis.

Equipped with a more stable grip on the epistemology of empathy, i.e. on the kind of understanding it can provide, the obstacles it has to face, the constitutive limits, and the surprising possibilities that characterize it, future work could turn to the issue of empathy's broader value as a form of mutual understanding.

Much of the previous work in this area has been carried in the field of moral psychology and philosophy, where scholars investigate empathy's value as a moral tool. As we saw in *chapter I*, scholars tend to have conflicting views on empathy when it comes to matters of morality.

However, I believe that future researches need not focus on moral issues even though they might intersect some concerns investigated by those disciplines.

I will suggest one main research thread that I think will be particularly interesting to investigate in the future as it constitutes a surprisingly underexplored topic in empathy research. Waiting for better terms, I will call this research thread *patient-oriented* as opposed to *agent-oriented*.

Much of the previous work has tended to focus on the agent's side of empathy, i.e. on the empathizers. However, people are not merely agents of empathy but also *patients* of empathy, i.e. targets of empathy's epistemic concerns.

As we saw in *chapter IV*, the philosopher Adam Morton insightfully noted how we can be rightly blamed for too superficially understanding another's perspective (Morton 2017). This seems to suggest that empathy might be required in certain interpersonal dynamics by the *patient* – rather than the agent – of empathy. This suggestion seems to be avowed by Paul Bloom who – while criticizing empathy on the moral ground – recognizes empathy's capacity to improve the quality of intimate relationships (Bloom 2016) and by Jodi Halpern who claims that empathy can have a remarkably positive impact on the physician-patient relationship and help the latter feel less isolated in their illness (Halpern 2001).

An alternative but nevertheless complementary thread of research, mostly carried out in the field of anthropology, highlighted the flip of the coin of what just said (Hollan 2008, 2018). Empathy can be sometimes regarded as an *intrusive* way of understanding others. By requiring the empathizer to place

themselves into the shoes of someone else, empathy runs the risk of being perceived as an undue usurpation of somebody else's perspective (see also Goldie 2011).

Especially when the cultural or interpersonal gap between the empathizer and the target is perceived as too wide, empathy is not only hostage of inaccuracy – i.e. the empathizer *actually* gets things wrong –, but also of *perceived* inaccuracy – i.e. the target feels like the empathizer does not get how they feel.

Focusing on the receiver's end of the empathic relation could help us illuminate underappreciated aspects of it. Instead of asking questions such as: What kind of understanding can empathy achieve? How well can empathizers understand their targets? Does empathy morally and epistemically improve empathizers? We could ask questions such as: Does empathy make people feel more understood? Is empathy something people require from others? Is empathy regarded as an "intrusive" way of understanding?

I firmly believe that shifting the focus from an agent-oriented to a patient-oriented view on empathy could open up new, underexplored paths of research and, in turn, help us to shed new light on our general understanding of empathy.

Bibliography

Addis, D. R. (2018). Are episodic memories special? On the sameness of remembered and imagined event simulation. *Journal of the Royal Society of New Zealand*, *48*(2–3), 64–88.

Anscombe, G. E. M. (2000). *Intention*. Cambridge, Mass.: Harvard University Press.

Aronowitz, S., & Lombrozo, T. (2020). Experiential Explanation. *Topics in Cognitive Science*, *12*(4), 1321–1336.

Avramides, A. (2020). Other Minds. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2020/entries/other-minds/. Accessed 18 January 2021

Bailey, O. (2018). Empathy and testimonial trust. *Royal Institute of Philosophy Supplement*, *84*, 139–160.

Balog, K. (2020). Either/Or: Subjectivity, Objectivity and Value. In J. Schwenkler & E. Lambert (Eds.), *Becoming Someone New: Essays on Transformative Experience, Choice, and Change*. Oxford, UK: Oxford University Press.

Barlassina, L., & Gordon, R. M. (2017). Folk Psychology as Mental Simulation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2017/entries/folkpsych-simulation/. Accessed 18 January 2021

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23.

Barrett, L. F., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2014). A psychological construction account of emotion regulation and dysregulation: The role of situated conceptualizations. In *Handbook of emotion regulation, 2nd ed* (pp. 447–465). New York, NY, US: The Guilford Press.

Batson, C., Early, S., & Salvarani, G. (1997). Perspective Taking: Imagining How Another Feels Versus Imaging How You Would Feel.

Bayne, T., & Montague, M. (2014). *Cognitive phenomenology*.

Bermúdez, J. L. (2017). *Understanding "I": Language and Thought*. Oxford, New York: Oxford University Press.

Bloom, P. (2017). Empathy and Its Discontents. *Trends in Cognitive Sciences*, *21*(1), 24–31.

Bloom, P. (2018). *Against empathy: the case for rational compassion*.

Boiten, F. A., Frijda, N. H., & Wientjes, C. J. (1994). Emotions and respiratory patterns: review and critical analysis. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *17*(2), 103–128.

Borkovec, T. D., & Hennings, B. L. (1978). The role of physiological attention-focusing in the relaxation treatment of sleep disturbance, general tension, and specific stress reaction. *Behaviour Research and Therapy*, *16*(1), 7–19.

Boven, L. V., & Loewenstein, G. (2003). Social Projection of Transient Drive States. *Personality & social psychology bulletin*.

Boven, V., Loewenstein, G., Dunning, D., & Nordgren, L. (2013). THREE Changing Places : A Dual Judgment Model of Empathy Gaps in Emotional Perspective Taking Leaf. /paper/THREE-Changing-Places-%3A-A-Dual-Judgment-Model-of-in-Boven-Loewenstein/9a3986d493c0c24c12c6802d854ec056f58f3b73. Accessed 27 January 2021

Boyle, A. (2020). The Impure Phenomenology of Episodic Memory. *Mind and Language*, *35*(5), 641–660.

C. Daniel Batson. (2009). These Things Called Empathy: Eight Related but Distinct Phenomena.

Camp, E. (2017). Perspectives in Imaginative Engagement with Fiction. *Philosophical Perspectives*, *31*(1), 73–102.

Campbell, J. (2015). L. A. Paul's Transformative Experience. *Philosophy and Phenomenological Research*, *91*(3), 787–793.

Campos, J. J., Campos, R. G., & Barrett, K. C. (1989). Emergent themes in the study of emotional development and emotion regulation. *Developmental Psychology*, *25*(3), 394–402.

Cappelen, H., & Dever, J. (2013). *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person*. Oxford University Press.

Cappelen, H., & Dever, J. (2017). Empathy and Transformative Experience Without the First Person Point of View. *Inquiry: An Interdisciplinary Journal of Philosophy*, *60*(3), 315–336.

Carruthers, P. (2018). Valence and Value. *Philosophy and Phenomenological Research*, *97*(3), 658–680.

Carruthers, P. (2020). *Human and Animal Minds: the Consciousness Questions Laid to Rest.* Oxford: Oxford University Press.

Charland, L. (2005). The Heat of Emotion: Valence and the Demarcation Problem. *Journal of Consciousness Studies*, *12*(8–10), 82–102.

Chartrand, T., & Bargh, J. (1999). The chameleon effect: the perception-behavior link and social interaction. *Journal of personality and social psychology*.

Coetzee, J. M. (2016). *The lives of animals* (Vol. 43). Princeton: Princeton University Press.

Colombetti, G., & Harrison, N.(2018) From physiology to experience: Enriching existing conceptions of "arousal" in affective science. In M. Tsakiris & H. D. Preester (Eds.), *The Interoceptive Mind: From Homeostasis to Awareness*. Oxford: Oxford University Press, 245–258.

Coplan, A. (2011). Understanding Empathy: Its Features and Effects. In A. Coplan & P. Goldie (Eds.) *Empathy: Philosophical and Psychological Perspectives*, Oxford: Oxford University Press, 3 -18.

Corns, J. (2014). Unpleasantness, Motivational Oomph, and Painfulness. *Mind and Language*, *29*(2), 238–254.

Currie, G. (2004). *Arts and minds*. Oxford: Clarendon Press.

Currie, G. (2020). *Imagining and Knowing: The Shape of Fiction*. Oxford: Oxford University Press.

Currie, G., & Ravenscroft, I. (2002). *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford: Oxford University Press.

Currie, G., & Ravenscroft, I. (2011). *Recreative minds: imagination in philosophy and psychology*. Oxford: Clarendon Press.

Cushman, F. (2020). Rationalization is Rational. *Behavioral and Brain Sciences*, *43*, 1–69.

Damasio, A. R. (1994). Descartes' Error: Emotion, Rationality and the Human Brain. New York*: Putnam*, *352*.

Darwall, S. (1998). Empathy, Sympathy, Care. *Philosophical Studies*, *89*(2–3), 261–282.

Davidson, D. (1963). Actions, Reasons, and Causes. *Journal of Philosophy*, *60*(23), 685.

Davies, H., Wolz, I., Leppanen, J., Fernandez-Aranda, F., Schmidt, U., & Tchanturia, K. (2016). Facial expression to emotional stimuli in non-psychotic disorders: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *64*, 252–271.

De Vignemont, F. Jacob. P. (2017). Vicarious Experiences Perception, mirroring or imagination? In J. Kiverstein (Ed.), *The Routledge handbook of philosophy of the social mind*. London; New York: Routledge, Taylor & Francis Group.

Decety, J., & Sommerville, J. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends in Cognitive Sciences,* 7(12), 527-533.

DeLancey, C. (n.d.). *Passionate Engines: What Emotions Reveal about the Mind and Artificial Intelligence*. *Passionate Engines*. Oxford: Oxford University Press.

Dennett, D. (1987). *The Intentional Stance*. Cambridge, Mass: MIT Press.

Deonna, J., & Teroni, F. (2012). *The Emotions: A Philosophical Introduction*. New York: Routledge.

Döring, S. A. (2003). Explaining Action by Emotion. *Philosophical Quarterly*, *53*(211), 214–230.

Döring, S. A. (2007). Seeing What to Do: Affective Perception and Rational Motivation. *Dialectica*, *61*(3), 363–394.

Eliot, T. S. (1957). *Essays on poetry and poets*. London; Boston: Faber and Faber.

Epley, N. (2008). Solving the (Real) Other Minds Problem. *Social and Personality Psychology Compass, 2*(3), 1455–1474.

Epley, N., Keysar, B., Boven, L. V., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of personality and social psychology*, 87(3), 327 -39.

Esch, T., Fricchione, G. L., & Stefano, G. B. (2003). The therapeutic use of the relaxation response in stress-related diseases. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, *9*(2), 23-34.

Eyal, T., Steffel, M., & Epley, N. (2018). Perspective Mistaking: Accurately Understanding the Mind of Another Requires Getting Perspective, Not Taking Perspective. *Journal of personality and social psychology*, 114(4), 547-571.

Fiske, S. T., & Taylor, S. E. (2020). *Social cognition: from brains to culture*. London: Sage Publication Ltd.

Galinsky, A., & Moskowitz, G. (2000). Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of personality and social psychology*, 78(4), 708 - 724.

Gendler, T. S. (2008). Alief in Action (and Reaction). *Mind and Language*, *23*(5), 552–585.

Gendler, T. S., & Kovakovich, K. (2006). Genuine Rational Fictional Emotions. In M. Kieran (Ed.), *Contemporary Debates in Aesthetics and the Philosophy of Art* (pp. 241–253). Blackwell.

Gendler, T. S., & Liao, S. (2016). The Problem of Imaginative Resistance. In J. Gibson & N. Carroll (Eds.), *The Routledge Companion to Philosophy of Literature* (pp. 405–418). Routledge.

Gilbert, D., & Wilson, T. (2007). Prospection: Experiencing the Future. *Science*, 317(5843):1351-4.

Gilbert, Daniel. (2006). *Stumbling on Happiness*. London: Vintage Books.

Gilovich, T., Kruger, J., & Medvec, V. (2002). The Spotlight Effect Revisited: Overestimating the Manifest Variability of Our Actions and Appearance. *Journal of Experimental Social Psychology*, 38(1), 93 - 99.

Goldie, P. (2002). Emotions, Feelings and Intentionality. *Phenomenology and the Cognitive Sciences*, *1*(3), 235–254.

Goldie, P. (2004). *On Personality*. New York: Routledge.

Goldie, P. (2009). *The emotions: a philosophical exploration*. Oxford: Clarendon Press.

Goldie, P. (2011). Anti-Empathy. In A. Coplan & P. Goldie (Eds.), *Empathy: Philosophical and Psychological Perspectives* (pp. 302 - 317). Oxford: Oxford University Press.

Goldman, A. (2006). Imagination and Simulation in Audience Responses to Fiction. In S. Nichols (Ed.), *The Architecture of the Imagination: New Essays on Pretence, Possibility, and Fiction* (pp. 41 - 56). Oxford: Calderon Press.

Goldman, A. I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.

Grimm, S. R. (2016). How Understanding People Differs from Understanding the Natural World. *Philosophical Issues*, *26*(1), 209–225.

Gross, J. J. (1998). The Emerging Field of Emotion Regulation: An Integrative Review. *Review of General Psychology*, *2*(3), 271–299.

Gross, J. J. (2015). Emotion Regulation: Current Status and Future Prospects. *Psychological Inquiry*, *26*(1), 1–26.

Halpern (2001) *From detached concern to empathy. Humanizing medical practice*. Oxford: Oxford University Press.

Hassabis, D., & Maguire, E. A. (2009). The construction system of the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1263–1271.

Hatzimoysis, A. (2007). The case against unconscious emotions. *Analysis*, *67*(296), 292–299.

Heal, J. (2003). *Mind, reason, and imagination: selected essays in philosophy of mind and language*. Cambridge: Cambridge University Press.

Hickok, G. (2014). *The Myth of Mirror Neurons: The Real Neuroscience of Communication and Cognition*. New York; London: W. W. Norton & Company.

Hoffman, M. L. (2007). *Empathy and moral development: implications for caring and justice*. New York: Cambridge University Press.

Hofmann, S. G., & Kashdan, T. B. (2010). The Affective Style Questionnaire: Development and psychometric properties. *Journal of Psychopathology and Behavioral Assessment*, *32*(2), 255–263.

Hollan, D. (2008) "Being There: On the Imaginative Aspects of Understanding Others and Being Understood", *Ethos*, 36, 475-489.

Hollan, D. (2018) Empathy across cultures. In: Maibom, H. L. (Ed.). *The Routledge handbook of philosophy of empathy* (pp. 341 - 353). New York: Routledge.

Hopkins, R. (2010). Imagination and Affective Response. In J. Webber (Ed.), *Reading Sartre: On Phenomenology and Existentialism* (pp. 100–117). New York: Routledge.

Hursthouse, R. (1991). Arational Actions. *Journal of Philosophy*, *88*(2), 57–68.

Ickes, W. J. (2003). *Everyday mind reading: understanding what other people think and feel*. Amherst, N.Y.: Prometheus Books.

Ismael, J. (2019). Experience, Transformation, and Imagination. *Rivista internazionale di Filosofia e Psicologia*, *10*(3), 330–338.

Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly*, *32*(April), 127–136.

Jacobson, E. (1928). Differential relaxation during reading, writing and other activities as tested by the knee-jerk. *American Journal of Physiology-Legacy Content*, *86*(3), 675–693.

James, W. (1884). What is an Emotion? *Mind*, *9*(34), 188–205.

James, W. (1890/1950). *The Principles of Psychology*. New York: Dover Publications.

Johnston, M. (2001). The Authority of Affect. *Philosophy and Phenomenological Research*, *63*(1), 181–214.

Kendler, K. S., & Campbell, J. (2014). Expanding the domain of the understandable in psychiatric illness: an updating of the Jasperian framework of explanation and understanding. *Psychological Medicine*, *44*(1), 1–7.

Kenny, A. (1963). *Action, Emotion And Will*. New York: Humanities Press.

Keysar, B., Lin, S., & Barr, D. (2003). Limits on theory of mind use in adults. *Cognition*, 89(1), 25 - 41.

Kind, A. (2016). Imagining Under Constraints. In P. Kung & A. Kind (Eds.), *Knowledge Through Imagination* (pp. 145 - 159). Oxford: Oxford University Press.

Kind, Amy. (2001). Putting the Image Back in Imagination. *Philosophy and Phenomenological Research*, *62*(1), 85–110.

Kind, Amy. (2016). *The Routledge Handbook of the Philosophy of Imagination*. New York: Routledge.

Koole, S. L. (2009). The psychology of emotion regulation: An integrative review. *Cognition and Emotion*, *23*(1), 4–41.

Korsgaard, C. M. (2008). *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*). New York: Oxford University Press.

Kovach, A., & Lancey, C. D. (2005). On Emotions and the Explanation of Behavior. *Noûs*, *39*(1), 106–22.

Kriegel, U. (2014). Towards a New Feeling Theory of Emotion. *European Journal of Philosophy*, *22*(3), 420–442.

Krueger, J., & Clement, R. W. (1997). Estimates of Social Consensus by Majorities and Minorities: The Case for Social Projection. *Personality and Social Psychology Review*, 1(4), 299 - 313.

Ledoux, J. (1984). Cognition and emotion: Processing functions and brain systems. In: M. Gazzaniga (Ed.) *Handbook of cognitive neuroscience* (pp. 357 - 368). New York: Plenum Press.

Leeuwen, N. V. (2016). Imagination and Action. In Amy Kind (Ed.), *The Routledge Handbook of Philosophy of Imagination* (pp. 286–299). New York: Routledge.

Leslie, A. M. (1994). Pretending and Believing: Issues in the Theory of ToMM. *Cognition*, *50*(1–3), 211–238.

Lewis, D. K. (1990). What Experience Teaches. In W. G. Lycan (Ed.), *Mind and Cognition* (pp. 29–57). Oxford: Blackwell.

Lewis, M. B. (2012). Exploring the positive and negative implications of facial feedback. *Emotion.*, *12*(4), 852–859.

Liao, S., & Gendler, T. (2019). Imagination. *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition).

Lim, N. (2016). Cultural differences in emotion: differences in emotional arousal level between the East and the West. *Integrative Medicine Research*, *5*(2), 105–109.

Loewenstein, G., & Boven, L. V. (2000). Egocentric Empathy Gaps between Owners and Buyers: Misperceptions of the Endowment Effect. In *Journal of personality and social psychology*, 79(1), 66 - 76.

Mahr, J. B. (2020). The dimensions of episodic simulation. *Cognition*, *196*, 104085.

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

Maibom, H. L. (2020). *Empathy*. New York: Routledge.

Matravers, D. (2017). *Empathy*. Cambridge: Polity.

McGinn, C. (2005). *Mindsight: Image, Dream, Meaning*. Cambridge (Ma): Harvard University Press.

McNeil, B. J., Pauker, S. G., Sox Jr, H. C., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *New England journal of medicine*, *306*(21), 1259–1262.

McRae, K., & Gross, J. J. (2020). Emotion regulation. *Emotion*, *20*(1), 1–9.

Michael, J. (2014). Towards a Consensus About the Role of Empathy in Interpersonal Understanding. *Topoi*, *33*(1), 157–172.

Millar, A. (2004). *Understanding People: Normativity and Rationalizing Explanation*. Oxford: Oxford University Press UK.

Moran, R. (1994). The Expression of Feeling in Imagination. *Philosophical Review*, *103*(1), 75–106.

Moran, R. (2011). Cavell on outsiders and others. *Revue internationale de philosophie*, (2), 239–254.

Morewedge, C. K., Gilbert, D., & Wilson, T. (2005). The Least Likely of Times. *Psychological science*, 16(8), 626 - 630.

Morton, A. (2017). Empathy and Imagination. In H. L. Maibom (Ed.), *The Routledge handbook of philosophy of empathy* (pp. 180–189). New York: Routledge.

Naber, M., Pashkam, M. V., & Nakayama, K. (2013). Unintended imitation affects success in a competitive game. *Proceedings of the National Academy of Sciences*, 110 (50), 20046-50

Nagel, T. (1974). What is It Like to Be a Bat? *Philosophical Review*, *83*(October), 435–50.

Nichols, S. (2008). Imagination and the I. *Mind and Language*, *23*(5), 518–535.

Nichols, Shaun. (2020). The propositional imagination. In *The Routledge companion to philosophy of psychology* (pp. 360–369). New York: Routledge, Taylor & Francis Group.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, *9*(3), 184–211.

Nussbaum, M. C. (1992). *Love's knowledge: Essays on philosophy and literature*. New York: Oxford University Press.

Nussbaum, M. C. (2009). *Upheavals of thought the intelligence of emotions*. Cambridge: Cambridge University Press.

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences*, *18*(11), 566–570.

Paul, L. A. (2014). *Transformative Experience*. Oxford: Oxford University Press.

Paul, L. A. (2015a). Transformative Experience: Replies to Pettigrew, Barnes and Campbell. *Philosophy and Phenomenological Research*, *91*(3), 794–813.

Paul, L. A. (2015b). Transformative Choice: Discussion and Replies. *Res Philosophica*, *92*(2), 473–545.

Paul, L. A. (2015c). Précis of Transformative Experience. *Philosophy and Phenomenological Research*, *91*(3), 760–765.

Paul, L. A. (2017). First Personal Modes of Presentation and the Structure of Empathy. *Inquiry: An Interdisciplinary Journal of Philosophy*, *60*(3), 189–207.

Peacocke, A. (2020). How literature expands your imagination. *Philosophy and Phenomenological Research*, 00, 1 - 22.

Pessoa, F., & Zenith, R. (2002). *The book of Disquiet*. London: Penguin.

Pitt, D. (2004). The Phenomenology of Cognition, Or, What Is It Like to Think That P? *Philosophy and Phenomenological Research*, *69*(1), 1–36.

Prinz, J. J. (2004). *Gut Reactions: A Perceptual Theory of the Emotions*. Oxford: Oxford University Press.

Prinz, J. J. (2011). Is Empathy Necessary for Morality? In: A. Coplan & P. Goldie *Empathy: Philosophical and Psychological Perspectives* (pp. 211–229), Oxford: Oxford University Press.

Railton, P. (2020). Rationalization of Emotion is Also Rational. *Behavioral and Brain Sciences*, *43:* e43.

Rausch, S. M., Gramling, S. E., & Auerbach, S. M. (2006). Effects of a single session of large-group meditation and progressive muscle relaxation training on stress reduction, reactivity, and recovery. *International Journal of Stress Management*, *13*(3), 273–290.

Recanati, F. (2007). *Perspectival Thought: A Plea for Moderate Relativism*. Oxford: Oxford University Press.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology.* New York, NY, England: Mcgraw-Hill Book Company.

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nature Neuroscience*.

Ruby, P., & Decety, J. (2004). How Would You Feel versus How Do You Think She Would Feel? A Neuroimaging Study of Perspective-Taking with Social Emotions. *Journal of Cognitive Neuroscience*, 4(5), 546-550.

Samson, D., Apperly, I., Braithwaite, J., Andrews, B. J., & Scott, S. E. B. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of experimental psychology. Human perception and performance,* 36(5), 1255 - 66.

Savitsky, K., Epley, N., & Gilovich, T. (2001). Do others judge us as harshly as we think? Overestimating the impact of our failures, shortcomings, and mishaps. *Journal of personality and social psychology*, 81(1), 44 - 56.

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). New York: Oxford University Press.

Scarantino, A., & Nielsen, M. (2015). Voodoo Dolls and Angry Lions: How Emotions Explain Arational Actions. *Philosophical Studies*, *172*(11), 2975–2998.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature reviews neuroscience*, *8*(9), 657–661.

Shannon, S. (2018). Do You See What I See? How Social Differences Influence Mindreading. *Synthese*, *195*(9), 4009–4030.

Shanton, K., & Goldman, A. (2010). Simulation theory. *Wiley Interdisciplinary Reviews. Cognitive Science*, *1*(4), 527–538.

Slote, M. A. & Oxford University Press. (2013). *Moral sentimentalism*. Oxford [i pozostałe: Oxford University Press.

Smith, A., & Hanley, R. P. (2009). *The theory of moral sentiments*. New York: Penguin Books.

Spaulding, S. (2016). Simulation Theory. In Amy Kind (Ed.), *The Routledge Hanbook of Philosophy of Imagination* (pp. 262–273). New York: Routledge.

Spaulding, S. (2017). Cognitive empathy. In H. Maibom (Ed.) *The Routledge Hanbook of Philosophy of Empathy* (pp. 13–21). New York: Routledge.

Spaulding, S. (2018). *How We Understand Others: Philosophy and Social Cognition*. New York: Routledge.

Sperber, D., & Mercier, H. (2017). *The Enigma of Reason*. Cambridge, Mass: Harvard University Press.

Spreng, R. N. (2013). Examining the role of memory in social cognition. *Frontiers in psychology*, *4*, 437.

Steinberg, J. (2014). An epistemic case for empathy. *Pacific Philosophical Quarterly*, *95*(1), 47–71.

Stueber, K. R. (2006). *Rediscovering empathy: agency, folk psychology, and the human sciences*. Cambridge, Mass.: MIT Press.

Stueber, K. R. (2011). Imagination, Empathy, and Moral Deliberation: The Case of Imaginative Resistence. *Southern Journal of Philosophy*, *49*(s1), 156–180.

Stueber, K. R. (2012). Varieties of Empathy, Neuroscience and the Narrative Challenge to the Contemporary Theory of Mind Debate. *Emotion Reviews*, *4*(1), 55–63.

Stueber, K. R. (2017). Empathy and understanding reasons. In *The Routledge handbook of philosophy of empathy*.

Ta, V. P., & Ickes, W. (2017). Empathic accuracy. In H. Maibom (Ed.) *The Routledge Hanbook of Philosophy of Empathy* (pp. 353–363). New York: Routledge.

Tamir, D. I., & Mitchell, J. (2010). Neural correlates of anchoring-and-adjustment during mentalizing. *Proceedings of the National Academy of Sciences*, 107(24), 10827-32.

Tamir, M., & Ford, B. Q. (2012). When feeling bad is expected to be good: emotion regulation and outcome expectancies in social conflicts. *Emotion*, *12*(4), 807–816.

Tamir, M., Mitchell, C., & Gross, J. J. (2008). Hedonic and Instrumental Motives in Anger Regulation. *Psychological Science*, *19*(4), 324–328.

Teroni, F. (2007). Emotions and Formal Objects. *Dialectica*, *61*(3), 395–415.

Teroni, F. (2018). Emotionally Charged: The Puzzle of Affective Valence. In F. Teroni, C. Tappolet, & A. K. Ziv (Eds.), *Philosophical Perspectives on Negative Emotions: Shadows of the Soul*. New York: Routledge.

Todd, A. R., Bodenhausen, G. V., Richeson, J. A., & Galinsky, A. D. (2011). Perspective taking combats automatic expressions of racial bias. *Journal of Personality and Social Psychology*, *100*(6), 1027–1042.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*, 519–545.

Tull, M. T., & Aldao, A. (2015). Editorial overview: New directions in the science of emotion regulation. *Current Opinion in Psychology*, *3*, iv–x.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-31.

Vignemont, F. de, & Jacob, P. (2012). What Is It Like to Feel Another's Pain? *Philosophy of Science*, *79*(2), 295–316.

Walton, K. (2015). Empathy, Imagination, and Phenomenal Concepts. In K. Walton (Ed.), *In Other Shoes: Music, Metaphor, Empathy, Existence* (pp. 1–16). Oxford: Oxford University Press.

Walton, K. L. (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge, Mass: Harvard University Press.

Weatherson, B. (2004). Morality, Fiction, and Possibility. *Philosophers' Imprint*, *4*, 1–27.

What Can We Learn From Taking Another's Perspective? (2018). In *Philosophical Perspectives on Empathy* (pp. 74–90). New York: Routledge.

Whiting, D. (2018). Consciousness and Emotion. In R. J. Gennaro (Ed.), *Routledge Handbook of Consciousness*. New York: Routledge.

Williams, B. (1973). Imagination and the Self. In *Problems of the Self* (pp. 26–45). Cambridge: Cambridge University Press.

Wirling, Y. (2014). Imagining Oneself Being Someone Else: The Role of the Self in the Shoes of Another. *Journal of Consciousness Studies*, *21*(9–10), 205–225.

Zajonc, R. (1984). On the primacy of affect. *American Psychologist*, 39, 117 - 123.