



UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

INTENTION, INTENTIONAL ACTION, AND SELF-KNOWLEDGE

A Transparency Account of Self-Knowledge of Intention

Davood Bahjat Fumani (Iran)

Thèse de doctorat présentée
à la Faculté des lettres et des sciences humaines
de l'Université de Fribourg (Suisse)

Approuvé par la Faculté des lettres et des sciences humaines
sur proposition des professeurs Gianfranco Soldati (premier rapporteur), Amir Saemi
(deuxième rapporteur), Johannes Roessler (troisième rapporteur)

Fribourg, le 18 décembre 2023
Le Doyen Professor Dominik Schöbi

To my mother, who sacrificed so much to ensure that I could pursue my dreams.

تقدیم به مادرم که فداکاری‌ها کرد تا بتوانم رویاهایم را دنبال کنم.

Table of Contents

Acknowledgements	7
Abstract	10
Introduction	11
1. Assumptions of Transparency Thesis	13
Abstract:	14
1.1. Introduction	15
1.1.1. Self-Knowledge and its problem	15
1.1.1.1. It starts with Descartes	15
1.1.1.2. Continues with Cartesians	18
1.1.1.3. Results in the doctrine of privileged access	21
1.1.2. Transparency Account	23
1.2. Moran's Account of Transparency	26
1.2.1. Introduction	26
1.2.2. Moran's reading of the doctrine of privileged access	27
1.2.3. The Difference Between the Deliberative and Theoretical Stance	30
1.2.4. Adopting a deliberative stance	31
1.2.5. Self-knower as an Agent: Moran's Transparency Account	34
1.3. Boyle's Account of Transparency	36
1.3.1. Introduction	36
1.3.2. The problem of Moran's account of self-knowledge and The Minimal Condition on Self-Knowledge	38
1.3.3. MCS demands representation of one's own condition as being of a certain kind	39
1.3.4. MCS Demands self-representation	41
1.3.5. What is the Ability to Understand	42
1.3.6. The one who has the ability to understand, knows one's own mental state	43
1.4. Byrne's Account of Transparency	45

1.4.1. Introduction	45
1.4.2. The method of transparency and its challenge	46
1.4.3. Byrne's Transparency Account	47
1.4.4. Doxastic Schema is self-verifying	48
2. Intentional Actions	51
Abstract	52
2.1. Introduction	53
2.2. Intentional Actions	54
2.3. Rationalist theory of Action	57
2.3.1. Anscombe's account of intentional action	57
2.3.2. One impediment and PTA's solution	59
2.4. Phronetic Theory of Action	61
2.5. Intentic Theory of Action	65
2.5.1. Cognitivism about intention (doxastic account of action)	66
2.5.1.1. One scenario, two approaches (1)	67
2.5.1.2. Two confusions in the cognitivists' approach and their possible response to clear up the confusions	68
2.5.1.3. At least a partial belief in success is necessary to do something intentionally	69
2.5.1.4. How PTA addresses the cognitivists' cases	70
2.5.1.5. The cognitivist's attempt to reject ITA fails	72
2.5.2. Consequentialist and Moral Theories of Action	73
2.5.2.1. One scenario, two approaches (2)	74
2.5.2.2. Two confusions in the consequentialists' approach and their possible response to clear up the confusions	74
2.5.2.3. How PTA addresses the consequentialists' cases	76
3. One's Knowledge of One's own Intentional Action	78
Abstract	79
3.1. Introduction	80
3.1.1. Why are we beginning with Anscombe? What is the starting point?	80
3.1.2. What is the challenge and how do we address it?	82

3.2. What one is doing intentionally, one is doing knowingly!	84
3.2.1 This knowledge is non-observational	84
3.2.2. It is hard to accept that the knowledge is non-observational	86
3.3. Two-Factor View	88
3.4. One's Knowledge of One's own Intentional Action is Practical Knowledge	90
3.5. How is such non-observational knowledge possible?	93
3.6. Setiya's account and PTA	97
4. Three Versions of the Transparency Account of Intention	99
Abstract	100
4.1. Introduction	101
4.2. Moran	103
4.2.1. Introduction	103
4.2.2. Moran's Transparency account of Intention	104
4.2.3. Objections	107
4.3. Boyle	112
4.3.1. Introduction	112
4.3.2. Boyle's Transparency account of Intention	113
4.3.3. Objections	115
4.4. Byrne	120
4.4.1. Introduction	120
4.4.2. Byrne's Transparency account of Intention	121
4.4.3. Objections	123
5. Phronetic Rule of Transparency	126
Abstract	127
5.1. Introduction	128
5.2. Non-observational but Inferential	130
5.3. The Transparency of Belief	132
5.4. The Transparency of Intention	135
5.5. Phronetic rule of transparency	140
5.5.1. In Search of a Transparent Rational Capacity	142

5.5.2. Practical Knowledge is not Reducible	145
5.5.3. Intention in Omission	146
5.5.4. Prospective Intentions	149
5.5.5. Transparency Transition	151
6. Conclusions & Outlooks	155
Abstract	156
6.1. Introduction	157
6.2. Conclusion	159
6.3. Why "Why?" not "What?"	160
6.4. Can we extend the Phronetic Rule of Transparency of Intention to other mental states: An idea for future research	164
Bibliography	167

Acknowledgements

Completing a PhD dissertation is a challenging task, and I am grateful to the many individuals who have supported and inspired me along the way. While it is impossible to name everyone, I want to thank a few of them for their unwavering encouragement and guidance.

First and foremost, I extend my heartfelt gratitude to my supervisor, Gianfranco Soldati, whose compassionate guidance, insightful comments, and stimulating support throughout my PhD journey have been nothing short of exemplary. Even before I arrived in Switzerland, Gianfranco was instrumental in helping me navigate the challenges of the visa application process and provided me with support and guidance from afar during my time in Iran. Beyond being a brilliant philosopher, Gianfranco is a great man whom I will always consider as an inspiration for what it means to be a truly exceptional human being. His profound influence on my way of thinking, be it philosophical or non-philosophical, is immeasurable. I consider myself privileged to have had the opportunity to work with him and learn from him.

I would like to express my gratitude to my colleagues and members of the Experience and Reason research group in Fribourg, whose presence has made my time there an intellectually stimulating and personally enriching experience. In particular, I am grateful to Thomas Aeppli, Frederic Auderset, Ralf Bader, Emmanuel Baierlé, Jiri Benovsky, Philipp Blum, Davide Bordini, Sharon Casu, Patrik Engisch, Christoph Haffter, Maté Herner, Isabel Kaesling, Elodie Malbois, Angela Martin, Maude Ouellette-Dubé, Mario Schärli, Samet Sulejmanoski, and Fabrice Teroni for their insightful feedback, constructive criticism, inspiring conversations, and the memorable moments of conviviality and joy that made the journey all the more enjoyable. I am particularly indebted to Donnchadh O'Conaill for the many constructive discussions that helped me to refine my ideas. I am also indebted to Martine Nida-Rümelin, whose comments on my presentations have been invaluable in polishing my ideas and sharpening my arguments.

I also want to express my deep appreciation to my close colleagues Miloud Belkoniene, Julien Bugnon, and Andrea Giananti for providing me with their

unconditional support from the beginning of my PhD. Miloud and Julien not only helped me with my philosophical investigation but also aided me in navigating the challenges of settling into a new country and academic community. They played an instrumental role in helping me feel integrated into Swiss society and the academic group. Additionally, sharing an office with Andrea and working on the same project provided me with an invaluable opportunity to present my ideas to him and seek his insightful input. He was always generous with his time, and our dialogues proved immensely beneficial in refining and developing my thoughts.

I consider myself fortunate to have been part of a collaborative project between the University of Fribourg and the University of Warwick (UK). This project allowed us to organise conferences held at both universities, where I had the opportunity to exchange ideas with and learn from Warwick members of the project, Johannes Roessler and Giulia Luvisotto. Johannes provided me with insightful feedback and great encouragement, demonstrating a genuine willingness to help me improve my ideas. I also enjoyed interacting with other Warwick's philosophical community members, notably Lucy Campbell and Guy Longworth, who offered invaluable insights and feedback that helped me sharpen my arguments. I would like to express my sincere gratitude to all of these individuals for their invaluable contributions to my academic journey.

I am grateful for the assistance and guidance provided by Iranian philosophers Laleh Ghadakpour, Amir Saemi, and Sajed Tayebi, who not only helped organise a conference in Iran but also provided valuable feedback and support throughout my academic journey. Additionally, I would like to express my appreciation to my Iranian friends in Fribourg, whose companionship and discussions enabled me to stay connected to my culture and homeland despite being far away therefrom: Mahshid Alizadeh, Farhad Babaei, Fatemeh Ehsannia, Aboubakr Moradi, Golnar Narimani, Jamshid Rajabi, Saeed Sadeghi, and specially Kiarash Ghassemi.

I express my deepest gratitude to my family, particularly my mother, Shaheen joon, and my brothers, Alireza, Arash, and Roozbeh. Their unconditional support and encouragement have enabled me to pursue my academic dreams abroad. Despite not being able to be with them physically for the past few years, I always felt their steady presence and support in every moment of my academic journey.

Finally, as I look back on my PhD journey, I realise that it would not have been possible without the endless love, patience, and encouragement of my wife, Yi Yang. She has been my constant source of support, believing in me when I doubted myself

and cheering me on through the ups and downs of academic life. Her unending patience and belief in me have sustained me through even the toughest moments. I am so grateful for her presence in my life and cannot imagine having gone through this journey without her presence by my side.

This dissertation was made possible by a grant from the Swiss National Science Foundation (SNSF) for the project "Perception, Rationality, and Self-Knowledge".

Abstract

This dissertation aims to elucidate the privileged access one has to one's own intentions by introducing a novel Transparency account of self-knowledge of intention. The study demonstrates that there is not only a close connection between knowledge of intentional action and self-knowledge but also that, in the case of intention, self-knowledge can be derived from knowledge of intentional action. To achieve this, the study presents a new theory of action called the Phronetic Theory of Action (PTA), which highlights the importance of knowing how to perform an action and exercising that knowledge in identifying intentional actions. According to this theory, an agent performs an action intentionally if they possess the knowledge of how to execute the action and exercise that knowledge. Based on this account, knowledge of intentional action, as non-observational knowledge, is explained. Next, this thesis delves into various accounts of transparency to demonstrate that PTA has the potential to provide a transparency account for self-knowledge of intention. The proposed Transparency account posits that knowledge of intention is transparent to knowledge of intentional action. In line with Setiya, this account contends that individuals with specific rational capacities can achieve self-knowledge by adhering to a rule of transparency. For self-knowledge of intention, this account proposes that knowledge-how is the specific rational capacity that one must utilise and suggests the Phronetic Rule of Transparency: If one has the capacity to know how to ϕ , and based on this capacity knows that one is ϕ -ing, one can, simultaneously and through the same capacity, ascribe the intention to ϕ to oneself. In conclusion, the study offers two suggestions. First, Anscombe's "why?" question could be replaced with the "what?" question that seeks the knowledge-how exercised by the agent. Second, drawing on Agentalist and Expressivist accounts of self-knowledge, it may be possible to extend the Phronetic Rule of Transparency for Intention to other mental states. By doing so, this dissertation provides a comprehensive framework for understanding the privileged access one has to one's own intentions and further explores the implications of this framework for self-knowledge in general.

Introduction

This dissertation aims to explain how one has privileged access to one's own intention. In contemporary debates about the philosophy of mind and epistemology, self-knowledge is of interest primarily because of the prevailing conviction that *one has privileged access to one's own mental states*, including one's beliefs, desires, and intentions, in the sense that our access to our own attitudes, as compared to the way we would have access to others' attitude, is more *epistemically secure (authoritative)* and more *directly shaped (immediate)*. The former feature usually refers to the conviction that one's report of one's own mental state is less fallible and less open to correction than one's report of others' mental attitudes. The latter feature refers to the conviction that one's knowledge of one's own mental state is non-inferential and not based on evidence or observation.

The inquiry into our unique access to our mental attitudes dates back to Descartes' claim that our mind is better known than our body. Since then, various theories have emerged to explain how we acquire self-knowledge. For instance, Inner Sense and Acquaintance accounts are two well-known theories that attempt to explain how we gain self-knowledge. The Inner Sense account posits that we gain self-knowledge by introspectively examining our own mental states. Acquaintance accounts hold that we have self-knowledge by being directly acquainted with our mental states. In this context, acquaintance refers to a direct, non-inferential, and non-conceptual relation between an individual and the individual's mental states.

This study focuses on another theory that has emerged as a critical view in recent years: the Transparency account. This account claims that knowledge of one's mental states is transparent to the world. The main idea is that one knows one's own mental state not by attending inward but by turning one's attention outward to the world. (Byrne 2018: 3). Among the philosophers who have been essential in the development of the Transparency account, Richard Moran, Mathew Boyle, and Alex Byrne have played prominent roles over the past two decades.

In this dissertation, we will show that Moran, Boyle, and Byrne's accounts of transparency with respect to the self-knowledge of intention are implausible. We will explain that what makes their account implausible is their negligence of the nature of intentional action and knowledge of one's own intentional action. By presenting a new

account of intentional action and knowledge of intentional action founded on 'knowledge-how', we will present our transparency account of self-knowledge of intention.

This dissertation is divided into six main chapters:

In the first chapter, we will briefly introduce the unique features of self-knowledge and explain what the problem of self-knowledge is. We will explain the Transparency accounts of self-knowledge and review the literature on Transparency accounts in the works of Richard Moran, Mathew Boyle, and Alex Byrne.

In the second chapter, we will present a theory of intentional action, which we call the Phronetic Theory of Action (PTA). In this theory, knowledge of how to do something and the exercise of this knowledge plays a crucial role in identifying intentional actions. According to this theory, an agent does an action intentionally if she knows how to do the action and exercise this knowledge. It is important to have an account of intentional action in this research because, on the basis of this account, we can explain knowledge of intentional action—which we will do in chapter three—and, through this explanation, we become able to address the problem of self-knowledge, particularly in the case of knowledge of one's own intention. Moreover, in chapter four, we will show how ignoring the true nature of intentional action and knowledge of one's own intentional action makes Moran, Boyle, and Byrne's accounts of transparency implausible.

In the third chapter, we will show that knowledge of one's own intentional action is distinct, and we will explain how such distinct knowledge is possible.

In chapter four we will assess Moran, Boyle, and Byrne's versions of the transparency account of intention, and show that they do not provide an appropriate epistemology for intention.

In chapter five, we will present a version of the transparency account of intention based on PTA. We will show that knowledge-how not only plays a crucial role in knowing one's own intentional actions, but also it can explain the privileged access that one has to one's own intention.

In chapter six, we will raise the question of whether we can extend the Phronetic Rule of Transparency of Intention to other mental states. We will point out that based on Agentalist and Expressivist accounts of self-knowledge, there might be a way to do it.

1. Assumptions of Transparency

Thesis

[M]y own present thinking, in contrast to the thinking of others, is transparent in the sense that I cannot distinguish the question “Do I think that P?” from a question in which there is no essential reference to myself or my belief, namely “Is it the case that P?” This does not of course mean that the correct answers to these two questions must be the same; only I cannot distinguish them, for in giving my answer to the question “Do I think that P?” I also give my answer, more or less tentative, to the question “Is it the case that P?”

(Edgley 1969: 90)

Abstract:

The objective of this chapter is to conduct a comprehensive review of the literature on transparency accounts of self-knowledge, with a particular emphasis on the contributions of Richard Moran, Mathew Boyle, and Alex Byrne. The chapter is structured into four main sections. The first section of this chapter provides a brief overview of the unique characteristics of self-knowledge and the challenges that it poses. The transparency account of self-knowledge is introduced and the reasons for selecting Moran, Boyle, and Byrne for review are explained. The second section of the chapter is devoted to Moran's account of transparency. This section analyses the crucial role played by Moran in developing the transparency approach, and a detailed review of his influential work, 'Authority and Estrangement: An Essay on Self-Knowledge', is provided. The third section focuses on Boyle's account of transparency. Two main criticisms leveled against Moran's account are outlined, and how Boyle addresses these objections is described. The main aim of this section is to provide a clear picture of Boyle's description of the minimal condition for self-knowledge, as well as to demonstrate how he defends Moran's account against criticisms by invoking this condition. In the fourth and final section, Byrne's inferential account of transparency is examined. Specifically, the reasons why Byrne believes that self-knowledge should be grounded in an inference from a premise about the external world to a conclusion about one's own mental states are analysed.

1.1. Introduction

1.1.1. Self-Knowledge and its problem

The main objective of this section is to provide an account of the emergence of literature on self-knowledge and to elucidate the problem of self-knowledge that philosophers aim to resolve. Along with this, it aims to provide a standard definition of some key terms that are commonly used in the literature on self-knowledge, such as infallibility, incorrigibility, immediacy, and others. The section is divided into three subsections.

The first subsection begins with an exploration of Descartes' ideas about knowledge of minds and the significant impact that his views have had on the literature on self-knowledge. In the second subsection, several theses that are derived from Descartes' insights are presented. The final subsection focuses on the dominant doctrine of privileged access that these theses constitute, which leads to the problem of self-knowledge.

1.1.1.1. It starts with Descartes

In recent years, there has been an increasing interest in the debate concerning self-knowledge. In philosophy, the term 'self-knowledge' is sometimes equated with knowledge of the self and its nature, whereas it can also be used to refer to knowledge of one's own mental states¹, including one's beliefs, desires, and intentions. Throughout this dissertation, the term 'self-knowledge' will be used in the latter sense. In the indicated sense, the debate regarding self-knowledge has its roots in a gap that Descartes exposed between mind (intelligent substance) and body (corporeal substance), and his well-known claim that our mind is better known than the body.

¹ In this dissertation, the terms 'self-knowledge' and 'knowledge of one's own mental states' are used interchangeably.

According to him, knowledge of our mind is prior to, and more certain and clear than knowledge of our body. (Descartes 1982: 6).

Descartes' thesis on knowledge of the mind is relevant to our discussion of knowledge of mental states because of Descartes' particular account of the mind. According to him, a mind is nothing but "a thinking thing" instantiated in different modes of thought,² including doubting, understanding, affirming, denying, willing, imagining, and sensing. (Descartes 1996: 18-19). Therefore, what he calls judgments about one's own mind are, in fact, what modern literature would call judgments about one's own mental states. Given this identity, we should not understand Descartes' cogito as a judgment about one's own mind in general (what he calls the intelligent substance) which is opposed to the body (the corporeal substance), but "a belief about one of our mental states"—I *believe* that I am thinking, then I exist. (Fernández 2013: 24; Burge 1996: 92; Borgoni 2018: 681). Although we do not intend to review Descartes' dualism, a brief explanation of these characteristics that Descartes ascribed to the knowledge of the mind—priority, certainty, and clarity—would help us understand the characteristics attributed to self-knowledge later.

Despite the fact that priority, certainty, and clarity refer to different characteristics in Descartes' epistemology, all three are in close relation to each other. In particular, with respect to his account of the knowledge of our mind, the discovery of each of these characteristics leads to the discovery of another.

Priority: Concerning the priority that knowledge of minds enjoys over knowledge of the body, Descartes maintains that if my mind makes judgment about something other than my mind, my judgment "gives even greater support for the judgement that my mind exists." (Descartes 1985: 196). As he says, it is possible that what my mind is judging about does not exist at all, but it is impossible that my judging mind does not exist itself. This priority that knowledge of our mind enjoys, according to Descartes, brings more certainty to this knowledge; as he puts it: "nothing can cause us to know something other than our mind, without at the same time bringing us with much more *certainty* to a knowledge of our mind itself." (Descartes 1982: 7). In what follows, we explain what this "certainty" is that Descartes talks about.

Certainty: In Descartes' view, *Knowledge* is a "certain cognition," but how he employs the term *certainty* varies. (Descartes 1985: 10). Although, for Descartes, we can have absolute/perfect certainty, which "arises when we believe that it is wholly

² See (Cottingham 1993: 128), and (Smith 2015: 106)

impossible that something should be otherwise than we judge it to be”, certainty can also come in degrees. We can be more certain of one thing than another (what Descartes calls moral certainty) — certainty can be a relative concept. (Descartes 1985: 290).³ Nevertheless, we can say that, for Descartes, the debate on certainty is generally tied with *doubt*, and he defines knowledge and certainty in terms of doubt. If one is more doubtful about something, one is less certain, and vice versa. Accordingly, if we cannot have any reason for doubting, we can reach the most perfect certainty. He says:

“Now if this conviction is so firm that it is impossible for us ever to have any reason for doubting what we are convinced of, then there are no further questions for us to ask: we have everything that we could reasonably want . . . For the supposition which we are making here is of a conviction so firm that it is quite incapable of being destroyed; and such a conviction is clearly the same as the most perfect certainty.” (Descartes 1984: 103)

For Descartes, the knowledge of our mind is more certain than the knowledge of our body because if we can doubt our judgments about bodies, we can never doubt our judgments about our own minds. He considers such a doubt impossible because “we cannot doubt them unless we think of them; but we cannot think of them without at the same time believing they are true, as was supposed. Hence we cannot doubt them without at the same time believing they are true; that is, we can never doubt them.” (Descartes 1984: 104).

Clarity: But how can one be firmly convinced of something and have no reason to doubt it? Descartes answers: because one has a clear and distinct perception of what one is convinced of. He says:

I am certain that I am a thinking thing; do I not therefore also know what is required for my being certain about anything? In this first discovery, there is simply a clear and distinct perception of what I am asserting; this would not be enough to make me certain of the matter if it could ever turn out that what I perceived with such clarity and distinctness was false. So I now seem to be able to lay it down as a general rule that whatever I very clearly and distinctly perceive is true. (Descartes 1996: 24)

In this respect, for Descartes, the knowledge of our mind is more certain than the knowledge of our body because of the clarity and distinction that the perception of the mind enjoys. In his view, clarity is in opposition to obscurity, and “if there is any certainty to be had, . . . it occurs in the clear perceptions of the intellect and nowhere else.” (Descartes 1984: 104). Descartes employs the verb ‘to perceive’ “for the purely

³ To read more on Certainty in Descartes, see: (Smith 2015: 51-56)

mental apprehension of the intellect." (Cottingham 1993: 143). In Descartes' literature, perceptions differ from sensory perceptions, such as hearing and seeing. If through sensory perceptions one sees something with the head's eye when light is reflected or emitted from the object, in the way Descartes uses the term, perceptions are acquired by means of the mind's eye ('natural light' or 'intuition'⁴). (Brueckner 2003: 187; Coliva 2016: 53). In explaining what is a clear perception, Descartes explains: "I call 'clear' that perception which is present and manifest to an attentive mind: just as we say that we clearly see those things which are present to our intent eye and act upon it sufficiently strongly and manifestly." (Descartes 1982: 20). The knowledge of our mind, according to Descartes, is clearer than the knowledge of our body because while our judgments about bodies may be obscure or liable to error, the object of our mind is so present and manifest to the mind's eye that in perceiving them there is no obscurity. There is no error in our judgments about our own mind.

Now that we have explained why, from Descartes' point of view, the knowledge of our mind is prior, more certain, and clearer than the knowledge of our body, it is time to proceed further and see the results this vision has brought to the literature on self-knowledge. In the following subsection, we discuss what is called the aspects of the traditional "Cartesian" conception of self-knowledge. (Brueckner 2003). We will explain the "Cartesian" features of self-knowledge, including infallibility, incorrigibility, transparency, immediacy, and being observational.

1.1.1.2. Continues with Cartesians

In the previous subsection, we explained why the knowledge of our mind, according to Descartes, is prior, more certain, and clearer than the knowledge of our body. We mentioned that these characteristics that Descartes attributed to the knowledge of the mind would help us understand the characteristics attributed to the knowledge of the self subsequently. Taking into account what we said in the previous subsection about Descartes' point of view on knowledge of the mind, in this subsection we outline the "Cartesian" characteristics of self-knowledge.

Many philosophers have attributed to Descartes the idea that self-knowledge is

⁴ See (Cottingham 1993: 94-6)

infallible and incorrigible.

Infallibility: As we have pointed out, according to Descartes, our judgments about our mind are not only more certain and clearer than our judgments about our body, but they are also free from all doubt and obscurity. This degree of certainty and clarity leads Descartes to assert that our judgments about our own minds are free from error. He says, “some of these perceptions [that mind has] are so transparently clear and at the same time so simple that we cannot ever think of them without believing them to be true.” (Descartes 1984: 104). Accordingly, he rejects the possibility of an error where we manifestly possess perfect certainty and says: “... I am incapable of error in those cases where my understanding is transparently clear.” (Descartes 1996: 48).⁵ This view of Descartes is the cornerstone of the infallibility thesis. According to the infallibility thesis, if one forms the belief that one is in a certain mental state, then that belief is indubitable, in the sense that one can have no reason to doubt the truth of that belief rationally. (Fernández 2013: 24). To put it in simpler terms, we can formulate the infallibility thesis in the following way:

- **The infallibility thesis:** one’s judgments about one’s own mental states “cannot be mistaken” (Cassam 2014: 43) and we can say that they “are guaranteed to be correct” (Brueckner 2003: 187).

Incorrigibility: For Descartes, making different judgments about the same thing is impossible when we perceive it clearly and distinctly. He says: “In many matters, people’s judgments disagree with their perceptions; but if we never make any judgement except about things we clearly and distinctly perceive - a rule which I always keep as well as I can - then we shall be incapable of making different judgements at different times about the same thing.” (Descartes 1991: 194-5). Here we should consider that in Descartes, we have clear and distinct perceptions only regarding the objects of our mind. In this regard, it can be concluded that no one can have a clear and distinct perception of other mental states, and consequently, “no one else can have good grounds for correcting them.” (Cassam 2014: 43). It should be noted that since, according to Descartes, we perceive our mental states clearly and distinctly, we are unable to make different judgments at different times about the same mental

⁵ In the third meditation, Descartes says: “Now as far as ideas are concerned, provided they are considered solely in themselves and I do not refer them to anything else, they cannot strictly speaking be false; for whether it is a goat or a chimera that I am imagining, it is just as true that I imagine the former as the latter. As for the will and the emotions, here too one need not worry about falsity; for even if the things which I may desire are wicked or even non-existent, that does not make it any less true that I desire them.” (Descartes 1996: 26).

state. Therefore, in Descartes' view, our judgments about the same mental state cannot be corrected/changed by ourselves. This approach of Descartes is the basis of the incorrigibility thesis. According to the incorrigibility thesis, if one forms the belief that one is in a certain mental state, then no one could provide evidence to show that the belief in question is false. (Fernández 2013: 24). More simply, we can formulate the incorrigibility thesis as follows:

- **The Incorrigibility thesis:** one's judgments about one's own mental states cannot be corrected by others.

To better understand the incorrigibility thesis, consider that S desires to eat ice cream. According to the incorrigibility thesis, no one can have a good ground for correcting S's belief that 'I have the desire to eat ice cream.' As Cassam (2014) mentions, the infallibility thesis and the incorrigibility thesis are different because it could theoretically be the case that S' beliefs about S's own desires are not error-free, but no one else could have good grounds for correcting them.

Two other theses concerning self-knowledge attributed to Descartes are transparency and immediacy.

Transparency (Self-intimation or omniscience thesis) and immediacy (directness): According to the transparency thesis, we are aware of everything that occurs in our mind, and according to the immediacy thesis, this awareness is immediate. Immediate awareness of mental states can be understood in two ways. First, in knowing one's own mental states, there is no time gap between the mind and mental states. Second, in knowing one's own mental states, there is no medium between the mind and mental states. As Newman (2019) mentions, Descartes believes in the immediacy of self-knowledge in both senses. Regarding the transparency and immediacy of mental states, Descartes says:

As to the fact that there can be nothing in the mind, in so far as it is a thinking thing, of which it is not aware, this seems to me to be self-evident. For there is nothing that we can understand to be in the mind, regarded in this way, that is not a thought or dependent on a thought. If it were not a thought or dependent on a thought it would not belong to the mind qua thinking thing; and we cannot have any thought of which we are not aware at the very moment when it is in us. (Descartes 1984: 171)

And somewhere else, when he is defining thought, he says:

Thought. I use this term to include everything that is within us in such a way that we are immediately aware of it. Thus, all the operations of the will, the intellect, the imagination and the senses are thoughts. I say 'immediately' so as to exclude the

consequences of thoughts; a voluntary movement, for example, originates in a thought but is not itself a thought. (Descartes 1984: 113)

We mentioned earlier (sec 1.1.1.1.) that, in Descartes' view, a mind is nothing but a thinking thing instantiated in different modes of thinking. For Descartes, this nature of the mind makes it evident that whatever belongs to the mind is transparent and immediate in the aforementioned sense.

In the current literature on self-knowledge, some aspects of the traditional "Cartesian" conception of self-knowledge are considered implausible. Instead of listing the characteristics of self-knowledge, in recent years, it is more common to say that one has "privileged access" to one's own mental states. In the next subsection, we explain why some Cartesian features seem implausible (Fernández 2013: 24) and what constitutes the prevalent doctrine of privileged access.

1.1.1.3. Results in the doctrine of privileged access

In the previous subsection, we explained four features of self-knowledge that derive from Descartes' account of knowledge of the mind. This subsection briefly presents an objection regarding the infallibility, incorrigibility, and transparency thesis. We will explain why these theses in the aforementioned version are too strong (Audi 2011) and therefore seem implausible. Furthermore, we will see how philosophers explain the immediacy thesis in the current literature on self-knowledge. Finally, we will explain how a moderate version of the infallibility thesis, the incorrigibility thesis, and the immediacy thesis constitutes the prevailing doctrine of privileged access that poses the problem of self-knowledge.

Many philosophers have pointed out that the claim that our judgments about our mental states are infallible and incorrigible is implausible. (Fernández 2013: 24; Cassam 2014: 44) There is no guarantee that our beliefs about what we believe cannot be mistaken. In addition, others might have reasonable grounds for correcting our assertion about our own attitudes. For example, Sarah's belief that she wants to be a lawyer can be mistaken, or others might have grounds for correcting her assertion that she wants to be a lawyer. Nevertheless, we can say that a common understanding of self-knowledge holds that self-knowledge is *authoritative* in the sense that one's judgments about one's own mental states, in comparison with the other's judgments

about those mental states, are less fallible and less open to correction.

With regards to the transparency thesis, we can say that it suffers the same problem as the infallibility and incorrigibility thesis, i.e., it is too strong. Just as we can have false beliefs about many of our mental states, we can also fail to form beliefs about many of our mental states. One might fail to recognise one's feelings or beliefs.

As far as the immediacy thesis is concerned, the story is different. Philosophers are more in disagreement with each other about how to interpret it rather than denying it or saying that it is too strong. A prevalent understanding of the immediacy thesis is that our beliefs about our own attitudes are groundless in the sense that they are non-inferential and non-evidentially based. (Cassam 2011: 1; 2017: 724). About the immediacy of self-knowledge, Boghossian says:

In the case of others, I have no choice but to infer what they think from observations about what they do or say. In my own case, by contrast, inference is neither required nor relevant. Normally, I know what I think—what I believe, desire, hope or expect—without appeal to supplementary evidence. Even where such evidence is available I do not consult it. I know what I think directly. I do not defend my self-attributions; nor does it normally make sense to ask me to do so. (Boghossian 1989: 7)

We can say that beliefs about one's own mental states are generally not derived from an inference or justified on the basis of other beliefs that are taken as evidence.

In the current literature on self-knowledge, a modest reading of the infallibility thesis, the incorrigibility thesis, and the immediacy thesis forms the dominant conviction about self-knowledge. According to the prevalent conviction, contrary to others' mental attitudes, one has *more epistemically secure (authoritative)* and *more directly shaped (immediate)* access to one's own mental states. The former feature usually refers to the conviction that one's report of one's own mental state is less fallible and less open to correction than one's report of others' mental attitudes. The latter feature refers to the conviction that one's knowledge of one's own mental state is non-inferential and not based on evidence or observation.

For example, when Sarah is thirsty, she usually knows about it, and it is seldom the case that she is wrong about her feeling or needs to have some evidence for her belief. Compare this with the case in which John believes that Sarah is thirsty. The extent to which he may be wrong is noticeable or at least higher than the situation in which Sarah forms such a belief about herself. In addition, what leads John to this judgment about Sarah's feeling is the evidence he has. However, it seems that Sarah does not need the same evidence for her belief.

So far, we can say that one has privileged access to one's own mental states in the sense that: 1. when they self-attribute beliefs, desires, and other attitudes, this self-attribution is not normally mistaken, and open to challenge by others, 2. one's access to one's own mental states is direct or immediate in the sense that it is not based on evidence and it is non-inferential. (Cassam 2014: 45).

The main problem of self-knowledge arises from the need to provide an explanation for such privileged access. The question is this: how can one make authoritative judgments about one's own mental states when these judgments are not based on evidence, observation, or inference? There are different theories of self-knowledge regarding the privileged access one has to one's own mind. A theory of self-knowledge is an account of how one knows one's own mental state in such a privileged way. A satisfactory theory of self-knowledge should explain why one enjoys such authoritative and immediate access to one's own attitude and explain why it is exclusively first-personal.

In the next section we will introduce one of the theories of self-knowledge that has received increased attention in recent decades: The Transparency theory of self-knowledge. The notion of transparency, which we will discuss in the following sections, should be distinguished from the notion of transparency, which we have discussed in the previous and current subsections. The former notion of transparency refers to the transparency of mental states to the world, and the latter refers to the transparency of mental states to the mind. In the next section, we will briefly explain how the transparency account attempts to provide an explanation for the privileged access one has to one's own mind.

1.1.2. Transparency Account

In the previous section, we mentioned that one has privileged access to one's own mental states in the sense that one's access to one's own mind is authoritative and immediate. There are a number of accounts in self-knowledge attempting to explain how one has such privileged access to one's own mental attitudes. Transparency account is one of these views around which considerable literature has recently grown up.

According to the Transparency account, knowledge of one's mental states is

transparent to the world. The main claim is that one knows one's own mental state not by attending "inward to the content" of one's own mind, but by turning one's "attention outward" to the world. (Byrne 2018: 3). Way introduces the transparency procedure as follows:

The basic procedure is that we ask ourselves whether to have attitude M, and if we conclude that M is the attitude to have, we judge that we have it, and if we conclude that it's not, we judge that we don't. (Way 2007: 226)

This procedure has its origins in Gareth Evans' claim regarding the self-ascription of belief. According to him:

[I]n making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me "Do you think there is going to be a third world war?," I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question "Will there be a third world war?" (Evans 1982: 225)

Evans believes that one's own belief is transparent to the fact that the content of this belief is the case. We can say that Evans' insight is not limited only to beliefs. The proponents of the Transparency account believe that, in general, an individual can come to know that she has a mental attitude by considering a question about the content of that attitude. (Way 2007: 223). As Dorit Bar-On says:

If asked whether I am hoping or wishing that p, whether I prefer x to y, whether I am angry at or afraid of z, and so on, my attention would be directed at p, x and y, z, etc. For example, to say how I feel about an upcoming holiday, I would consider whether the holiday is likely to be fun. Asked whether I find my neighbour annoying, I would ponder her actions and render a verdict. (Bar-On 2004: 106)

Bar-On describes how one's attention is directed towards the object of a mental attitude, such as a belief, desire, or emotion when considering questions about that attitude. However, the Transparency approach has been subject to different objections. One of the main objections to the Transparency views of self-knowledge is that they have a limited scope. (Paul 2014: 301). One manifestation of this objection is the claim that these views cannot explain how we possess first-personally authoritative knowledge of attitudes we held prior to the question being raised:

Looking outward to determine whether one believes that p leads to the formation of a judgment about whether p, which one can then self-attribute. But use of this process does not explain one's access to judgments already in place. (Gertler 2011b: 126)

Another form of the objection is that the Transparency theories cannot be applied to mental states or events that are not sensitive to reason, such as pain or irrational

attitudes. We get more familiar with this objection in the following chapters.

Despite these objections, the last two decades have witnessed a growing trend towards Evans's transparency account, largely due to Richard Moran's efforts to elaborate on the concept of the Transparency method. In his influential book '*Authority and Estrangement: An Essay on Self-Knowledge*,' Richard Moran defends an account of Transparency that he calls "the fundamental form of self-knowledge⁶." (Moran 2001: 150). Moran emphasises in his account that a self-knower has particular first-person access to oneself because he makes up his own mind in the sense that he actively participates in forming his own attitudes. However, some philosophers cast doubt on this claim that his account is the fundamental form of self-knowledge. Their objection, in short, is that Moran's account includes some cases of self-knowledge in particular and is not expandable to all mental states one would have.

Unlike these philosophers, Mathew Boyle believes that Moran's claim regarding the fundamentality of his Transparency account is defensible. Boyle's Transparency account originated from the minimal condition that he considers for self-representation. According to him, someone only can represent one's own mental state, if one understands "a form of the first-person" and the "predicates that apply to the first-person." Having this minimal condition of self-knowledge in mind, Boyle tries to indicate that "recognizing one's power to make up one's mind in the way Moran describes" is the only way to understand a form of the first-person used in self-representation. (Boyle 2009: 153).

Both Moran and Boyle present a non-inferential account of Transparency. For, they believe that one's access to one's own mental states is immediate in the sense that one's access is not based on observation, evidence, or any inference. Nevertheless, Alex Byrne, another philosopher who has contributed to the literature, believes that the Transparency account must be based on "an inference from world to mind." (Byrne 2011: 203). Otherwise, according to Byrne, we cannot explain how transitioning from a certain fact 'p' to a mental state 'I believe p' would be possible.

The following three sections are dedicated to providing a detailed analysis of Moran, Boyle, and Byrne's understanding of self-knowledge, and how they aim to resolve the problem of self-knowledge through their Transparency accounts. Each of these sections focuses on reviewing the Transparency accounts presented by one of the philosophers.

⁶ Moran means that his account of self-knowledge includes all mental states in general.

1.2. Moran's Account of Transparency

1.2.1. Introduction

Richard Moran is credited with bringing the Transparency account to increasing attention in the literature on self-knowledge. He is among the first to elaborate on Evans' well-known insight that you can determine your mental attitude by gazing outward into the external world.

Moran agrees that the problem of self-knowledge is rooted in the immediate and authoritative access one has to one's own mental states. As he argues, a satisfactory account of self-knowledge should explain how such immediate and authoritative awareness of one's own mental states is possible and why only one's own mind is knowable in this way.

To present his own account, Moran first determines the scope of his account. He distinguishes between "two basic categories of psychological state" to which one can have such immediate and authoritative access: *occurrent states*⁷ and *standing attitudes*. (Moran 2001: 9-10; Cassam 2011: 2). Occurrent states include "sensations and passing thoughts" (Ibid.) are mental events. "[T]hinking a given thought as you think it" (Cassam 2017: 724) or having a given sense as you perceive something are some examples of occurrent states. By contrast, standing attitudes are states like belief, desire, and intention.⁸ They "aren't mental events, and you can have them even when you aren't actively entertaining them." (Ibid.).⁹ Moran makes it clear from the beginning that, in his account, he leaves out the occurrent states and focuses on the privileged access that one has to one's own standing attitudes. He says:

"I will have comparatively little to say here about the case of sensations, which I

⁷ Is different from what some philosophers like McLaughlin and Tye calls occurrent thought; "our conscious mental states of thinking that P". (McLaughlin and Tye 1998: 350)

⁸ Brie Gertler uses the notion of 'occurrent mental state' for belief and other attitudes, by which she means the occurrent mental state that we have currently. She says, "the method of transparency does not explain access to pre-existing occurrent beliefs, but rather produces new judgments, which we can then knowingly self-attribute." (Gertler 2011b).

⁹ According to Boyle, there are at least two fundamentally different types of self-knowledge: an active type, which involves standing attitudes, by which we know our own judgments, and a passive type, which involves occurrent states, by which we know our sensations. (Boyle 2009)

Regarding the passivity of occurrent states Cassam says: "Passing thoughts are passive in the sense that they are (i) not necessarily responsive to reason, and (ii) states from which one can distance or dissociate oneself." (Cassam 2011: 3)

believe raises issues for self-knowledge quite different from the case of attitudes of various kinds." (Moran 2001: 9-10).

Putting aside the occurrent states, Moran believes that the immediate and authoritative relation one has to one's standing attitudes cannot be explained through empirical approaches— in which one perceives one's own mental states through a perceptual faculty called 'inner sense,' which is somehow comparable to the ordinary perceptual faculties— and theoretical explanations which concentrate "on questions of belief and judgment as applied to some static realm of mental facts." (Moran 2001: 27). According to him, such theoretical stances toward oneself neglect the person's role as an epistemic agent who can link her mental attitudes and make up her mind about what she does or believes. (Moran 2001: 33&56). Moran contends that first-person ascriptions are not merely a "special claim to truth" that "stand without the benefit of evidence." Rather, they are a self-knower's answers to deliberative questions like "What am I to believe?". In answer to this kind of question, one should also satisfy the "rational demand" expected from a rational person, in the sense that one should be "able to subject one's attitudes to review in a way that makes a difference to what one's attitude is." (Moran 2001: 26&64). In other words, one has an active role in forming one's own attitude; as Moran describes it, one's stance towards oneself is *deliberative*.

According to Moran, the Transparency account is consistent with the active role of a self-knower and provides a suitable solution to the problem of self-knowledge. This is because one's own mental state is transparent to one's response to a deliberative question. The following subsections will provide an explanation of Moran's interpretation of self-knowledge and the problem it poses. This will be followed by a detailed account of his Transparency approach.

1.2.2. Moran's reading of the doctrine of privileged access

In his rewarding book '*Authority and Estrangement: An Essay on Self-Knowledge*,' Moran considers an asymmetry between one's knowledge of one's own attitudes and other people's attitudes. In accordance with the doctrine of privileged access (1.1.1.3.), he admits that self-knowledge is distinct from the knowledge one can have of others'

mental attitudes. He also agrees that the privileged access one has to one's own mental life is due to the immediate and authoritative access one has to one's own mental attitudes— self-knowledge enjoys a kind of immediacy and authority.

Moran's understanding of immediacy and authority is in line with what we explained earlier in section 1.1.1.3.. About these two features, he says:

“while a person may learn of someone else's beliefs or other attitudes from what she says and does, he may arrive at knowledge of his own attitudes in a way that is not based on evidence or observation of himself. In this sense, a person may know his own mind “immediately,” yet nonetheless declare his belief with an authority that is lacking in anyone else's observation-based description of him.” (Moran 2001: xxix)

As he explains, one's knowledge of one's own mental state is immediate in the sense that one's access to one's own mind is not based on evidence, observation of oneself, or any inference. With regards to the self-knowledge of belief, Moran holds that “we can see it as a rational requirement on belief, on being a believer, that one should have access to what one believes in a way that is radically non-evidential, that does not rely on inferences from anything inner or outer” (Moran 1997: 143). Some philosophers accept that one's knowledge of one's own mental state is non-observational and not based on evidence, while they believe that self-knowledge depends on a kind of inference. However, Moran's conviction is that self-knowledge is independent of any kind of inference. He says, “if introspective awareness is anything at all—that is, anything distinct from the knowledge of the mental life of others—then it seems it must be something different from any knowledge based on inference.” (Moran 2001: 17). In this idea, he follows Boghossian who states that we know what others think by inferring “from observations about what they do or say” but when it comes to our own thoughts, “inference is neither required nor relevant.” (Boghossian 1989: 7).

Furthermore, as mentioned earlier in 1.1.1.2., the immediacy thesis has its roots in Descartes' view of the knowledge of the mind. As we explained earlier, Descartes believes that one of the characteristics of immediate awareness is that there is more certainty about our thoughts than any other things, in the sense that our own thoughts are infallible. However, contrary to Cartesian's conviction, Moran believes that “immediacy per se does not confer any greater reliability or freedom from error on a set of judgments.” Therefore, he sees the main claim of the immediacy thesis not in connection with the infallibility thesis, but as “a wholly negative claim about the mode of first-person access, that is, awareness that is not inferred from anything more basic.”

(Moran 2001: 11).

With regard to the authority enjoyed by access to one's own mental states, Moran endorses a modest reading of the infallibility and incorrigibility theses and concedes that the strong reading of these theses could be rejected in certain cases. According to him, one's judgments about oneself, unlike the judgments about other's mental states, "enjoy a particular epistemic privilege" in the sense that "they are taken to have a good prima facie claim to truth which may be overruled only in special cases." (Moran 2001: 11). Moran does not take for granted that access to one's own mental attitudes is infallible. However, by attributing authoritative access to self-knowledge, he acknowledges a kind of "reliability" that one's judgments about oneself possess. According to Moran, this reliability is essentially or exclusively first-personal and cannot be extended to the access one has to others' minds.

For Moran, this sense of authority is in close connection with the authoritative role that, according to him, a self-knower plays in determining one's own mental state. Concerning this role, Moran says: "[T]he person can be seen as the author of the state of mind itself, in the sense of being the person who originates it and is responsible for it." (Moran 2001: 113) In this regard, for Moran, authority in self-knowledge emerges from the authoritative role one has towards the formation of one's mental states. This dimension of the authority of self-knowledge constitutes the central part of Moran's account.

So far, we have explained Moran's reading of the doctrine of privileged access. Given the immediacy and authority one enjoys in accessing one's own attitudes, Moran raises two questions:

"How is it possible for there to be knowledge of some contingent matter of fact (e.g., the facts about what I believe or hope for) that is not based on observation of some kind? And in what sense is this knowledge supposed to be essentially or exclusively first-personal?" (Moran 2001: 13).

We already got familiar with the first question, where we explained the main problem of self-knowledge in section 1.1.1.3.. There we said that the main problem of self-knowledge is: how can one make authoritative judgments about one's own mental life when these judgments are not based on evidence, observation, or inference? However, Moran also has another question regarding the privileged access that one has to one's own mental states. He also wonders why this authoritative access one has to oneself is exclusively first-personal and cannot be extended to the access one has to the minds of others. This question stems from the connection Moran makes

between having a first-person stance in self-knowledge and having privileged access. To answer these questions, Moran distinguishes between two different ways of explaining self-knowledge: deliberative stance and theoretical stance. He argues that only the deliberative stance enables us to explain the unique features of self-knowledge, i.e., immediacy and authority.

1.2.3. The Difference Between the Deliberative and Theoretical Stance

To address the problem of self-knowledge, Moran first examines the very question that one would form when one inquires about one's own mind; questions like, 'do I believe P?', 'what do I want?', and 'how do I feel?'. According to Moran, one's question about one's own mind could be answered in two different ways: *descriptive* and *normative*. The way one answers the question depends on the way one raises the question. Although questions of oneself can be expressed by the same interrogative sentence, like "what do I believe?", in the questioning of one's own belief, there can be two different readings of it. On the one hand, this question interrogates 'what is this that I believe?'. On the other hand, this question interrogates, 'what is this that I am to believe?' The former looks for a "descriptive" answer, while there is some "normative" expectation for answering the latter, in the sense that "the person himself plays a role in formulating how he thinks" and what he should believe. (Moran 2001: 58-59; 2004: 424)

When one is ignorant of an antecedent fact about oneself, for instance, his or her current emotional status, he or she would ask this question of himself or herself, 'how do I feel?' by which he or she means, 'what is this that I feel?'. It seems that our approach to answering this kind of question is similar to the way that, from a third-person perspective, one would try to address the question. A proper answer to this question is a descriptive answer by which one describe one's current feeling as a discovery of a fact of which one were ignorant. (Moran 2001: 58).

However, there is another kind of question concerning one's own mind that gets answered exclusively first-personal. According to Moran, the questions about one's own mental states that interest us here are of this type. For example, with respect to

intention, Moran says, one does not passively confront one's own attitude, but one has an active role in forming one's own intention by "making up" one's mind. (Moran 2001: 56). To answer the question "what am I going to do?" one needs to think carefully and make a reasonable decision about what one has to do.

Moran calls the first kind of questions "theoretical questions" and the second kind of questions "deliberative questions." Although these two questions may have a similar form, we answer them in two different ways. According to him, we adopt a *deliberative stance* towards ourselves when we answer a deliberative question, and we adopt a *theoretical stance* towards ourselves in answering a theoretical question.

1.2.4. Adopting a deliberative stance

According to Moran, a deliberative question about oneself "is answered by a decision or commitment of some sort, and it is not a response to ignorance of some antecedent fact about oneself." (Moran 2001: 58). Answering a question about the attitudes of others does not necessarily lead to any commitment or decision. Saying that someone else wants an ice cream doesn't make me committed to buying one. The same happens when one responds to a question about one's own attitudes by adopting a theoretical stance. However, according to Moran, adopting a deliberative stance toward our own attitudes can lead to a commitment to our own attitudes. Normally, when we want ice cream, we buy one. Concerning beliefs, when one questions one's attitude and adopts a deliberative stance by uttering 'I believe that p,' Moran suggests that the individual not only expresses a belief that p but also commits to the truth of p. According to Moran, "to be a believer at all is to be committed to the truth of various propositions." (Moran 1997: 147). Accordingly, when one takes a deliberative point of view, one's judgment that 'I believe that p' provides sufficient reasons to act in accordance with p's truth. In contrast, when one adopts a theoretical standpoint, the belief that p does not provide "any reasons for acting on p's truth." (Brueckner 2003: 196). This is analogous to the attitudes one may have towards somebody else's belief. When one adopts a theoretical stance in answering the question 'what do I believe?', "there is a gap between one's judging that one believes that p ... and one's committing, in action, to the truth of p, just as there is a gap between one's judging that another believes that p and one's committing, in action, to the truth of p." (Brueckner)

We mentioned earlier that with regard to the problem of self-knowledge, Moran was looking for an answer to this question: how can one make authoritative judgments about one's own mental attitudes, while such judgments do not rely on observation of some kind? And why is this authoritative access to oneself exclusively first-personal? According to Moran, to give a proper answer to this question, one must take into consideration the deliberative stance one usually takes toward one's own mental states:

The special features of first-person awareness¹⁰ cannot be understood by thinking of it purely in terms of epistemic access ... to a special realm to which only one person has entry. Rather, we must think of it in terms of the special responsibilities the person has in virtue of the mental life in question being his own. (Moran 2001: 32).

In the theoretical stance, one simply seeks evidence in order to find out the best answer regarding one's own mental attitude. However, as Moran argues, this stance is inappropriate for making judgments about one's own attitudes. According to him, one has a "deliberative role" in determining one's own mental state. We, as rational creatures, are responsible for our attitudes and usually evaluate and revise them in line with our reasons. For instance, with regard to our intentions, the question "what am I going to do?" cannot be answered by a theoretical inquiry. It is because "my knowing what I will do next is not based on evidence or other reasons to believe something, so much as it is based on what I see as reasons to do something." (Moran 2001: 56). One intends to do something based on one's decision, and one has an active role in forming one's own intention. With regard to the other mental states, according to Moran, one plays such an active role as well.

According to Moran, we not only have access to what we believe or intend, but we also form our beliefs and intentions, and we are responsible for our attitudes in the sense that we, as rational creatures, should revise our mental states if necessary. As he puts it, the authority one has over one's own attitudes is linked to the active role one plays in the formulation of those attitudes:

The primary thought gaining expression in the idea of 'first-person authority' may not be that the person himself must always 'know best' what he thinks about something, but rather that it is his business what he thinks about something, that it is up to him. In declaring his belief, he does not express himself as an expert witness to a realm of psychological fact, so much as he expresses his rational authority over that realm. (Moran 2001: 123-4).

¹⁰ He means the special features of self-knowledge.

Since what we think, believe, or intend is our own business, in response to questions like 'what do I believe?' it is our responsibility to form a rational attitude. According to Moran, in forming our rational attitudes, we direct our attention "at least equally towards the outward," towards the object of our response, as towards ourselves. (Moran 2001: 59). For Moran, turning our gaze outward places us in an agential/ first-person engagement with our own mental attitudes:

One is an agent with respect to one's attitudes insofar as one orients oneself toward the question of one's beliefs by reflecting on what's true, or orients oneself toward the question of one's desires by reflecting on what's worthwhile or diverting or satisfying [...]. There is a role for the agent here insofar as we may speak of a person's responsibility for his attitudes. (Moran 2001: 64).

When Moran talks about "turning the gaze outward," he is talking about reflecting on what is true, what is worth wanting, what is worth doing, and other such critical questions. Moran believes that directing one's attention outward to the world in order to form an attitude for which one is responsible is a crucial point for presenting a theory of self-knowledge. According to him, this agential/ first-person engagement one has with one's own attitudes is neglected in observational self-knowledge. For him, empirical approaches and theoretical explanations fail to address the problem of self-knowledge, since they try to tackle that problem by means of describing special observational access that one has to one's own attitudes. These methods fail to explain the deliberative stance one typically adopts toward one's own attitudes, "namely that one grasps one's attitude as revisable attitudes for which one is responsible." (Gertler 2011a: 174). These observation-based accounts of self-knowledge view the relationship between the individual and her mental attitudes as a passive relationship— not a revisable commitment for which they are responsible. In other words, it seems that we are alienated from our mental attitudes if we obtain them exclusively through observation, since we observe and grasp them with a distance from ourselves, like the observation and grasp that one can have of someone else's mental states from a third-person perspective.

So far, we have seen that, according to Moran, an individual usually adopts a deliberative stance towards herself. We have said that this means an individual has an active role in the formation of her own attitudes. Furthermore, we have said that in Moran's point of view, this active role can explain the privileged access that someone has to her mental states and why this access is exclusively first-personal. We have mentioned that having an active role means directing our attention outwards,

i.e. reflecting on what to think, what to do, or similar questions. We have said that, in Moran's opinion, this kind of thinking places us in an agential/first-personal engagement with our own mental attitudes in the sense that we form revisable attitudes for which we are responsible. Now one might ask how can adopting a deliberative stance explain why we have immediate and authoritative access to our own mental states?

1.2.5. Self-knower as an Agent: Moran's Transparency Account

We have said that, according to Moran, our agential role in forming our own mental states on the basis of our own deliberation can explain the immediate and authoritative access we have to our attitudes. The question now is how he advances his argument and makes the connection between deliberation and self-knowledge. We said that, for Moran, in response to the questions like 'what do I believe?' it is up to us to form a rational attitude, and in forming our rational attitudes, we direct our attention "towards the outward." In response to the questions like 'what do I believe?', we are not turning our gaze inward to find out our beliefs in a passive way. Instead, our answer to 'what do I believe?' is derived from our active attention to the world and forming our belief based on our evidence for it. In other words, Moran believes our mental attitudes are transparent to the world.

As Moran says, "with respect to belief, the claim of transparency is that from within the first-person perspective, I treat the question of my belief about P as equivalent to the question of the truth of P." (Moran 2001: 62-63). Thus, by having sufficient reasons for P, I am committed not only to the truth of P, but also to the truth of my belief about P. We can find the origin of Moran's idea about self-ascription of belief in Gareth Evans's approach that we had quoted earlier:

In making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward — upon the world. If someone asks me "Do you think there is going to be a third world war?", I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question "Will there be a third world war?" I get myself in a position to answer the question whether I believe that p by putting into operation whatever procedure I have for answering the

question whether p. (Evans 1982: 225)

For Moran, if one is a rational agent possessing a genuine system of beliefs, then one must adopt a deliberative stance upon one's own beliefs, from which the question 'Do I believe that p?' is transparent to the question 'Is p true?'. (Brueckner 2003: 196). Consider a situation where an individual believes that it is raining due to compelling evidence. The sound of rain falling on the window, the observation of cloudy weather, and raindrops on the window are the primary sources of evidence leading to the conclusion that it is currently raining. If asked about her belief, "Do you believe it is raining?" the individual can immediately respond, "Yes, I believe it is raining." Upon confirmation, she can attribute the belief to herself in an authoritative manner without requiring further evidence. As Moran says:

[I]nsofar as the person can answer the question about his belief in a way that conforms to transparency, we will have the beginnings of an explanation both of how the person can know his mind 'immediately,' without appeal to evidence, and how the answer arrived at has a kind of authority not shared by any other person's ascription to him of some state of mind. (Moran 2004: 424)

Accordingly, when asked whether one believes it is raining, the individual needs only to consider the question, "Is it raining?" If there is compelling evidence to support the belief that it is raining, the individual can confidently endorse the belief and state, "I believe that it is raining." Moran argues that one can be cognisant of her own attitude in this manner and ascribe it to herself immediately and authoritatively through a declaration. Moran refers to this sort of declaration of one's own attitude as an avowal:

'Avowal' is defined as a way of answering a question about one's belief or other attitude that obeys the 'Transparency Condition', hence a form of self-knowledge that is immediate because transparent to a corresponding question that is directed outward, upon the world. (Moran 2004: 424)

In Moran's point of view, one has privileged access to one's own mental attitude "only if one can avow the attitude", and it is only possible if one "can learn of the attitude by using the transparency method." (Gertler 2011a: 175). When one discovers the belief that it is raining using the transparency method, Moran suggests that one can express a properly first-personal apprehension of oneself by avowing, "I believe that it is raining."

1.3. Boyle's Account of Transparency

1.3.1. Introduction

It does not go wrong if we say that Mathew Boyle's account of self-knowledge has generally drawn on Moran's view of self-knowledge that we explained in the previous section. Following the popular viewpoint in the literature on self-knowledge, Boyle agrees that what makes self-knowledge different from the knowledge that one would have of other's minds is the privileged access to one's own mind. He also agrees that this privileged access is due to the immediacy and authority normally enjoyed by having access to one's own mind. Boyle's understanding of immediacy and authority is in harmony with what we explained earlier in section 1.1.1.3.. He outlines the immediacy involved in self-knowledge by saying that to ascribe a mental state to oneself, one does not need "the sorts of evidence that would be required for his ascription of such states to another person." With respect to the authority, he remarks that in attributing a mental state to oneself, self-ascriptions "are not normally liable to the same kinds of error that afflict ascriptions of such states to other people." (Boyle 2009: 136). For Boyle, too, the general problem of self-knowledge lies in the very existence of these two features in one's access to one's own mind, while one's knowledge of others' mental states is observational or inferential:

"The general problem of self-knowledge is to explain how we can be in a position to speak about our own minds in such an immediate and authoritative manner, while still counting as speaking about the very same states that can be known to others only on the basis of observation or inference." (Boyle 2009: 136)

Boyle praises Moran's approach to the problem of self-knowledge and argues that to solve the problem of self-knowledge, like Moran, we should take a different approach from what he calls the *epistemic approach*. According to Boyle, the epistemic approach encompasses any account that accepts that "in the normal, non-alienated case, being in a given mental state M and believing oneself to be in M are two distinct psychological conditions, and ... the task of a theory of self-knowledge is to explain how these conditions come to stand in a relation that makes the latter knowledge of the former."(Boyle 2011: 235). In Boyle's view, all accounts relying on inner sense, the causal relation between first-order mental states and second-order beliefs, or inference

adopt an epistemic approach. The primary purpose of such accounts is to indicate a particular way of arriving at a mental attitude.

Boyle contrasts the epistemic approach with the approach adopted by Moran to the problem of self-knowledge, which he calls the “reflective approach”. According to this approach, self-knowledge is a distinctive sort of knowledge, and we are mistaken if we think that we can conceive of self-knowledge in the same way that we conceive of knowledge of other objects, and just the epistemological details are special. (Boyle 2011: 239). Reflectivists, according to Boyle, offer an account of how self-knowledge is possible by presenting an account “that is primarily metaphysical rather than epistemological.” (Boyle 2011: 235). According to the reflectivists’ account that Boyle presents, the primary goal is not to indicate “a special way of arriving at” a mental attitude, but to indicate how normal knowledge of one’s own mental attitude “reflects something about what [having a mental attitude] is.”¹¹ (Ibid).

As we mentioned, Boyle's account of self-knowledge has inspired by Moran's view on self-knowledge. As we reviewed in the subsection 1.2.1., Moran makes a distinction between occurrent states and standing attitudes. We said that the scope of Moran's account of self-knowledge rules out occurrent states like sensations and passing thoughts. Thus, Moran's account does not explain the subject's access to all mental states and works only for those attitudes in which the subject has an active role in forming their content, i.e., standing attitudes. This is in contrast to Moran's claim that the self-knowledge he has described is “the fundamental form of self-knowledge.” (Moran 2001: 150). Many critics of Moran reject this claim based on the idea that a satisfactory account of self-knowledge must explain how one has knowledge of one's mental states in general. They are looking for a uniform account which does not cover only some particular mental states, like belief or desire.

To solve the issue, Boyle explores the minimal condition one needs to meet in order to know one’s own mental state. According to him, if someone knows her own mental state and can express it in sentences like “I'm in pain,” then she should understand the different parts that compose these sentences. Having this minimal condition of self-knowledge in mind, Boyle tries to indicate that “recognizing one's power to make up one's mind in the way Moran describes” is the only way that one can understand a form of the first-person used in self-representation. (Boyle 2009: 153). To represent Boyle’s transparency account, we will first describe the problem of Moran's account

¹¹ the words in the brackets are mine. In the original text is written: “reflects something about what believing is.”

that Boyle attempts to address. Then, we will show how Boyle tackles this problem by appealing to the minimal condition of self-knowledge.

1.3.2. The problem of Moran's account of self-knowledge and The Minimal Condition on Self-Knowledge

To solve the problem of self-knowledge — i.e., to explain how self-knowledge benefits from immediacy and authority — Boyle starts by assessing Moran's transparency account. In this regard, Boyle raises this question: given the acceptability of Moran's form of self-knowledge, "how much of our actual self-knowledge takes this form." (Boyle 2009: 138). In his book, Moran claimed that the self-knowledge he described is "the fundamental form of self-knowledge." (Moran 2001: 150). Nevertheless, there are some objections against Moran's claim about the fundamentality of his form of self-knowledge. Boyle counts two of these objections: first, his account excludes not only non-observational and non-deliberative knowledge that one has of sensations, but also such knowledge that one has of one's "appetites" (like groundless desires), as well as one's "recalcitrant attitudes" (like out of control feelings of anger). (Boyle 2009: 139). Second, even regarding the mental states like belief and desire for which Moran's account seems more defensible, one would be immediately and authoritatively aware of having such attitudes "without going through any process of conscious deliberation." (Boyle 2009: 139). Moran's account explains how one makes his mind by thinking about a state of affairs. However, if a belief is *already formed in the past*, one usually does not need to return to the process and make up one's mind again.

These two objections demonstrate that Moran's claim about the fundamentality of self-knowledge that he presents is contestable. This paves the way for proponents of what Boyle calls the "Uniformity Assumption" to reject Moran's transparency account. As Boyle says:

The assumption underlying these criticisms is evidently that we should seek some common explanation of all of the cases in which we can speak immediately and authoritatively about our own mental states. We could call this the Uniformity Assumption, for it amounts to the demand that a satisfactory account of our self-

knowledge should be fundamentally uniform, explaining all cases of "first-person authority" in the same basic way. (Boyle 2009: 141).

However, Boyle neither finds this assumption plausible nor thinks that by describing the fundamental form of self-knowledge, Moran means that his account of self-knowledge includes all mental states in general. In order to reject the Uniformity Assumption and defend a sense of "fundamental" on which Moran's claim is true, he appeals to what he calls a minimal condition of self-knowledge. Based on this minimal requirement, one's avowals can express one's knowledge of the states that one reports, only if one understands "whatever sentences one uses to express this knowledge." (Boyle 2009: 142).

- **The Minimal Condition on Self-Knowledge (MCS):** When one expresses the mental state that one is in, one must understand whatever sentences are used for this expression.

MCS suggests a link between expressing oneself and an ability to speak about one's own mental state. According to Boyle, while the assumption of using an articulate language is not a necessary condition for self-knowledge, using a verbal expression, for example, using an utterance like 'I'm in pain,' is a way in which one "mature language-users" can express oneself. (Boyle 2009: 148-9). Moreover, what Boyle has in mind is the traditional description of the problem of self-knowledge, which has been associated with "the ability to say, without observation or inference", what one's own mental state is. Hence, Boyle holds that a satisfactory account of self-knowledge should explain this ability. (Boyle 2009: 149).

According to MCS, when someone avows, "I'm in pain," she must understand this sentence—understanding such a sentence in general demands understanding its different elements in particular. Different elements by which "I'm in pain" is composed are the first-person pronoun "I" as the subject and "am in pain" as the predicate. With this in mind, Boyle begins to explain why one must understand these elements when ascribing a mental state to oneself.

1.3.3. MCS demands representation of one's own condition as being of a certain kind

As Boyle argues, in order to meet MCS, one needs to meet a further condition that is more basic: "self-knower must represent her own condition as being of a certain kind." (Boyle 2009: 143). Explaining how this representation of one's own certain condition must be, Boyle brings an example, showing the difference between an expression of mental states by "a mature competent speaker" and a child or an animal like a parrot. Consider a parrot who is taught to say, 'I'm in pain' when it feels pain. The verbal expression of pain by the parrot could not be counted as an expression of its knowledge of its own pain. Parrot's reaction to the pain by crying out, "I'm in pain," is an "automatic response," indicating it feels pain. However, there is no evidence to indicate that the parrot can distinguish the state of pain as a certain condition that it is in. As Boyle formulates, there is no "grasp that it is in pain," but it "is merely a learned addition to whatever repertoire of behaviors parrots naturally have for expressing pain." (Boyle 2009: 143). In other words, the parrot's expression "I'm in pain" is just a passive reaction that it is conditioned to have when it is in pain, and this expression does not show any understanding of its condition.

Boyle compares the parrot case to that of a mature, competent speaker who sincerely says she is in pain. In the latter situation, the subject can distinguish the condition of being in pain from all other conditions she might experience. According to Boyle, the mature, competent speaker understands what she says when expressing her pain. Therefore, the expression of pain in this case indicates not only the condition she is in but also her knowledge of being in that particular condition. As a result, Boyle distinguishes between two senses of expressions of a mental state:

[W]e must distinguish between two senses in which a kind of behavior might be said to "express" a mental state: the sense exemplified in the utterances of our imagined parrot, which we might call the manifestation sense (expression_M), and the sense exemplified in the superficially similar utterances of a competent speaker, which we might call the representation sense (expression_R). (Boyle 2009: 144).

By this distinction, Boyle intends to indicate that the expression of self-knowledge sometimes merely manifests the presence of a certain mental state, but does not represent the privileged access to that state. In other words, in expression_M , there is a kind of alienation between the subject and her own mental states - the kind of alienation that can be found in a third-person judgment of mental life. Whereas this type of alienation does not occur in expression_R , in which self-expression is a representation of one's own certain condition.

In summary, MCS demands that when one expresses_R one's mental state, one

understands what being in that certain mental state means. Consequently, we can reformulate the minimal condition on self-knowledge as below:

- **The Minimal Condition on Self-Knowledge* (MCS*):** When one expresses the mental states that one is in, one must understand whatever sentences are used for this expression, in the sense that the expression must be a *representation* of one's own *certain* condition.

1.3.4. MCS Demands self-representation

As explained above, Boyle told us that expression of self-knowledge should be representative of one's certain mental state. It explains why one must understand the predicate that one uses to ascribe a mental state to oneself. Then he moves on to discuss in more detail what he means by the subject's ability to represent *her own* mental state first-personally. Here, Boyle intends to explain why one must also understand the first-person pronoun used in ascribing a mental state to oneself. According to him, in order to represent one's own mental state, one must not only represent the particular mental state one is in, but also be able to attribute it to oneself and represent it as one's own mental state. Accordingly, Boyle describes the type of self-representation in question here as "the kind that a subject with the relevant linguistic abilities (1) would be able to report in (2) an utterance involving a form of the first person." (Boyle 2009: 148). Otherwise, if one cannot recognise oneself as "the object of the representation," one might represent the *actual state* one is in, but they do not manage to represent it as *one's own* state. (Boyle 2009: 148).

Thus, the utterance 'I'm in pain' is representational when one understands what *pain* as a certain kind of condition is. It is self-representational when one ascribes this certain kind of condition *to oneself* and identifies oneself as the person who is in pain (by using the first-person pronoun *I*).

Identifying oneself as a person in a certain condition, and using the first-person pronoun "I" in an utterance to indicate this recognition, demands the subject's understanding of the meaning of "I". In other words, one's usage of "I" should express a comprehending representation of oneself. Therefore, to be more precise about MCS, MCS*—which only required an understanding of what it means to be in that particular mental state—should be reworded to properly reflect the demand for an

understanding of the meaning of "I". We can reformulate MCS* as below:

- **The Minimal Condition on Self-Knowledge** (MCS**):** When someone expresses the mental state that one is in, one must understand whatever sentences are used for this expression, in the sense that one's expression must be a *representation* of one's own *certain* condition.

1.3.5. What is the Ability to Understand

After explaining the importance of understanding the sentences we use to express our mental state, Boyle delves into the factors that allow us to comprehend the content of our own utterances. According to Boyle, comprehending the content of an utterance requires that the subject be able to reflect on the relationships between the content of a given sentence and the content of various other sentences. In his words, "at least part of what is required is that the subject should be able to reflect on relationships between the content of any given sentence and the content of various other sentences" (Boyle 2009: 150). As Boyle puts it, we can demonstrate our understanding of a sentence by using different elements of that sentence in other sentences. For instance, in the case of understanding the first-person pronoun "I" in the sentence "I'm in pain," Boyle contends that if one cannot produce other sentences with "I" and understand them, it becomes evident that one does not comprehend the meaning of "I" in that sentence. The same holds for the predicate portion of the sentence, i.e., "am in pain."

- **The Ability to Understand (AU):** To meet MCS**, one must be able to reflect on relationships between the content of any given sentence about one's own mind and the content of various other sentences.

However, as noted by Boyle, this condition alone is insufficient for indicating that one comprehends the content of one's own utterance. He adds that one "must be able to recognize relationships between the truth of any one [namely, any given sentence] and the truth of others." (Boyle 2009: 150). Boyle considers this condition because a true recognition of a common element in different sentences depends on recognizing the relationships that exist between their contents and taking a stance on what is true based on those relationships. By recognizing these relationships, one can adjust one's utterance and takes "a stand on what is true, a stand related to various other stands she might take." (Boyle 2009: 150). In simple terms, the ability to compare different

sentences involves the ability to evaluate their truth, adjust a sentence and take a position on what is true. This ability is necessary for understanding the relationships between the contents of different sentences and, ultimately, for comprehending the sentences we use to represent our own mental states.

- **The Ability* to Understand (AU*):** To meet MCS**, one must be able to reflect on the relationships between the content of a given sentence about one's own mind and the content of various other sentences, so that one can recognise which sentences are true.

Boyle is already approaching Moran's view of transparency. This becomes more evident when Boyle argues that in order to exercise AU* and hold a given claim to be true, one must be able to reflect on the grounds for that claim. (Boyle 2009: 150-1). Boyle says: "a comprehending speaker must be able to entertain a certain sort of "Why?"-question about the claims she makes, a question that asks for grounds that show the claim in question to be true." (Boyle 2009: 151).

Accordingly, we can reformulate AU* in the form below:

- **The Ability** to Understand (AU**):** To meet MCS**, one must be able to reflect on the relationships between the content of a given sentence concerning one's own mind and the content of various other sentences, so as to be able to recognise, on the basis of one's grounds, which sentences are true.

After explaining the ability required to understand an utterance by which one expresses one's own mental state, Boyle then examines how the exercise of this ability allows them to self-attribute a mental state and claim self-knowledge.

1.3.6. The one who has the ability to understand, knows one's own mental state

According to Boyle, the one who has the ability to self-represent, in the manner he has described, is the one who is entitled to accompany the expression of one's own mental state with 'I believe'. The reason for this entitlement is that someone who sincerely expresses her mental state and has sufficient grounds to take her own claim true is a subject whose speech already expresses her beliefs: "when she (nondeceptively) says "p," she will be affirming something she takes to be true, and since to take something

to be true just is to believe it, she will also be entitled to say "I believe that p.'" (Boyle 2009: 151). In simple terms, Boyle wants to say that if one has the ability to understand whatever one claims and can give a sincere, true statement about one's own mental state, then the claim not only manifests one's mental state but also represents one's knowledge of one's mental state. To understand better, remember Moran's transparency method. We said that, according to Moran, for the one who adopts a deliberative stance upon one's own beliefs, the question 'Do I believe that p?' is transparent to the question 'Is p true?'. Boyle wants to say that Moran is correct because if one can sincerely claim that P is the case after deliberating whether P, one's understanding of one's claim entitles one to say, "I believe that P." A parrot does not have this entitlement because it cannot understand its claim in such a way that the truth of its claim rests on the reasons arrived at in a deliberation.

Explaining how one can know one's own mind, Boyle comes to the conclusion that if this is the right way to achieve such knowledge, then Moran is right to claim that his characterization of self-knowledge is the most fundamental:

If this is right, then we are in a position to say why the kind of self-knowledge that Moran characterizes is fundamental. It is fundamental because the ability to say what one believes in the way Moran specifies is intimately connected with the kinds of representational abilities that must be possessed by a subject who can make comprehending assertions. (Boyle 2009: 151).

In Boyle's view, Moran's characterization of self-knowledge is fundamental, for his account of self-knowledge rests on "a precondition of thought of oneself," namely, having a deliberative capacity to make up one's own mind and form one's own mental state in an active way. Therefore, for Boyle, those who insist on the Uniformity Assumption deny this necessary precondition and "thus undermine the possibility of self-knowledge in general." (Boyle 2009: 156).

1.4. Byrne's Account of Transparency

1.4.1. Introduction

Alex Byrne is one of the main philosophers who, in recent years, has contributed to the development of the transparency account. Similar to other philosophers such as Moran and Boyle, Byrne argues that the unique access individuals have to their own mental states is crucial for understanding self-knowledge. Byrne argues that self-knowledge is distinguished from other forms of knowledge by two features - privileged access and peculiar access. However, his definition of these two features aligns with the immediacy and authority aspects of self-knowledge discussed earlier in section 1.1.1.3., indicating that his understanding of self-knowledge is in harmony with the prevalent conviction presented earlier.

According to Byrne, privileged access refers to the more epistemically secure access one has to one's own mental attitudes. He argues that beliefs about one's own mental attitudes are less fallible compared to beliefs about the minds of others. This feature resembles the authority feature of self-knowledge.

Byrne uses the term "peculiar access" to refer to the more directly shaped belief one has about one's own mental states, which resembles the immediacy feature of self-knowledge. Byrne adds that the way in which one normally comes to know about one's own mental attitudes is exclusive to oneself, and this method cannot be used to gain knowledge about the minds of others. (Byrne 2011: 202). Although Byrne avoids explicitly using terms like "immediate," "non-evidential," or "non-observational" when explaining peculiar access, he does mention that one can know one's own mental state "without observing oneself at all," "without evidence," with "inadequate evidence," or "no evidence at all." (Byrne 2018: 8).

According to Byrne, the problem of self-knowledge lies in our privileged and peculiar access to our own mind, and the central problem "is to explain (or explain away) the privileged and peculiar access we enjoy to our mental states." (Byrne 2018: 16). Byrne adds that these two features are distinct aspects of self-knowledge, and neither entails the other. (Byrne 2005: 81; 2011: 202). Byrne acknowledges that there are different approaches regarding the characteristics of self-knowledge, with some

philosophers taking for granted the presence of one feature while rejecting the other. For instance, a behaviourist may believe in privileged access to one's own mind while rejecting any peculiar access. (Byrne 2005: 81). Although Byrne recognises the independence of these two features, he believes that a good theory of self-knowledge can reveal the connection between them.

While Richard Moran and Boyle argue that one's knowledge of one's own mind cannot be based on inference, Byrne maintains that one cannot know one's own mental state unless one arrives at such knowledge through an inference from the world to the mind. (Byrne 2011: 203). Accordingly, he believes that the Transparency thesis relies on an inference schema by which one can conclude that one possesses a particular mental state based on a premise about something in the external world. Byrne introduces different schemas for different mental states, and in this section, we will focus on the schema he uses to demonstrate the possession of belief, which he calls the Doxastic Schema.

1.4.2. The method of transparency and its challenge

Byrne believes that the Transparency account has the potential to solve the problem of self-knowledge and explain the connection between privileged and peculiar access to one's own mental states. He aligns with Evans's insight and understands the Transparency method as a process in which one knows about one's own mental states by turning one's attention outward to the world. (Byrne 2011: 203; 2018: 3).

As noted previously, after Evans, there have been significant discussions and analyses on the method of transparency by Moran, Boyle, and others. However, according to Byrne, they have failed to explain the very process of transitioning from the world to the mind. Byrne contends that what remains unclear in their account is how questions about one's own mental states are transparent to questions about the non-mental world. He highlights this gap by asking the question, "suppose that I examine the evidence and conclude that there will be a third world war. Now what?" (Byrne 2011: 203). For Byrne, the question is how and why one can move from knowledge about the external world - that there will be a third world war - to a conclusion about the mind - that one believes there will be one.

1.4.3. Byrne's Transparency Account

Byrne's response to the question of how such a transition occurs is that "the next step involves an inference from world to mind: I infer that I believe that there will be a third world war from the single premiss that there will be one." (Byrne 2011: 203). Unlike Moran and Boyle, who believe that self-knowledge is not based on any inference, Byrne finds it impossible to bridge the gap between the world and the mind without relying on an inference that enables the transition from a premise about the world to a conclusion about the mind.

Therefore, we can state Byrne's Rule of Transparency (BRT) as follows: The transition from world to mind should be based on an inference from a worldly premiss to a conclusion about one's own mind.

Although applying an inference would allow one to make this transition, it remains unclear "how can such an inference yield knowledge?" (Byrne 2011: 203). One not only has a mental attitude towards what is happening in the world, but also knows it. In what follows, we will explore how Byrne's inferential account of the Transparency view makes it possible to explain self-knowledge.

Byrne owes his idea of inferential Transparency to André Gallois' doxastic schema:

$$\frac{p}{\text{I believe that } p}$$

The doxastic schema suggests that if someone discovers that 'p,' then she can infer that 'I believe that p.' For example, if one recognises that 'it is raining,' one can infer that "I believe that it is raining." Gallois argues that the inference from 'p' to 'I believe that p' is standardly justified. If someone recognises 'p' as a fact, then she is entitled to infer that she believes p. Otherwise, it would be an instance of Moore-paradoxicality. As Moore argues, *in accordance with ordinary language* when one asserts something, one implies that one believes it. (Moore 1993: 210). According to Moore, "it is absurd or nonsensical to say such things as 'I don't believe it's raining, but as a matter of fact it is' or (what comes to the same thing) 'Though I don't believe it's raining, yet as a matter of fact it really is raining'." (Moore 1993: 207). In this regard, Gallois believes that inferences based on DoS are standardly warranted:

"It is Moore-paradoxical for me to allow that p, but deny that I believe p. If I allow that

p then I should allow that I believe p. So, if I allow p, I am entitled to infer that I believe p. Call any instance of the doxastic schema a Moore inference. The claim we are considering is that Moore inferences are standardly warranted." (Gallois 1996: 47).

It is paradoxical to acknowledge that 'p' is the case but deny that one believes it, as acknowledging the truth of 'p' implies that one believes it. Therefore, if someone acknowledges 'p,' she is entitled to infer that she believes it. Gallois believes that this inference is typically warranted and can be used to explain how one can arrive at knowledge of one's own beliefs based on external world facts.

However, both Byrne and Gallois acknowledge that the doxastic schema is not a good inference, as it "is neither deductively valid nor inductively strong." (Byrne 2011: 204). A person does not need to be aware of any inductive correlation between 'p' and 'I believe that p' in order to make the inference. Additionally, Gallois notes that the inference from 'p' to 'I believe that p' "is not validated as an example of inference to the best explanation."¹² (Gallois 1996: 47). The best explanation for 'I believe that p' is not the recognition of 'p'. These issues raise questions about whether the doxastic schema can solve the problem of self-knowledge of belief. It appears that the schema is not knowledge-conducive and cannot provide us with knowledge about our beliefs.

In short, the puzzle of transparency for belief is that: "how can one come to know that one believes that p by inference from the premise that p?" (Byrne 2018: 99). In the following section, we will explore how Byrne attempts to solve this puzzle.

1.4.4. Doxastic Schema is self-verifying

Byrne attempted to address the puzzle of Transparency by appealing to epistemic rules. An epistemic rule is a general principle or guideline that one can follow to acquire knowledge or justify beliefs. It provides a way to reason or evaluate evidence in order to determine "what it would be most rational to believe under various epistemic circumstances." (Boghossian 2008: 472). Byrne explains epistemic rules as conditionals of the following form: "If conditions C obtain, believe that p." (Byrne 2005: 94).

¹² Inference to the best explanation is a form of reasoning where one infers a hypothesis based on how well it explains the available evidence. The hypothesis is then accepted as true if it provides a better explanation of the evidence than any other alternative hypothesis. (Harman 1965: 89)

For example, when someone hears the doorbell ringing, she can usually arrive at the conclusion that there is someone at the door. In this case, she is following the epistemic rule that: "If the doorbell rings, believe that someone is at the door." (Ibid.)

According to Byrne, the transition from 'P' to 'I believe that p' in the doxastic schema can be explained in light of the following epistemic rule:

BEL If p, believe that you believe that p.

Byrne claims that BEL is an epistemic rule that allows us to move from 'p' to 'I believe that p' using the Doxastic Schema. However, this does not necessarily mean that BEL is a reliable inference method, as it can still produce a true conclusion even when the premise is false. Therefore, there is a question of how much we can trust BEL to give us accurate self-knowledge.

Byrne's response to this challenge is that the doxastic schema is "self-verifying," which means that when we use it to self-ascribe a belief, we end up with a true self-belief, regardless of whether the premise is true or false. In other words, the doxastic schema reliably yields safe beliefs, which are unlikely to be false. This makes BEL a reliable inference rule that gives us privileged self-knowledge. (Byrne 2011: 206; Paul 2015: 1533; Winokur 2021: 30).

Furthermore, Byrne argues that BEL delivers peculiar self-knowledge because it is only applicable to oneself and cannot be used to acquire knowledge about the beliefs of others. In other words, BEL provides a method for self-ascribing beliefs that is not available for ascribing beliefs to others. (Winokur 2021: 30). This means that the method of BEL yields peculiar self-knowledge that is exclusive to oneself. Therefore, the BEL rule provides a way for individuals to acquire self-knowledge that is not only privileged but also peculiar to themselves.

To conclude this section, Byrne's Transparency account of self-knowledge of belief provides an explanation for both privileged and peculiar access one has with regard to one's own beliefs. Privileged access is explained by the strong self-verifying nature of the doxastic schema. Peculiar access is explained because the method only works for oneself, as trying to infer that someone else believes p from the premise that p can often be misleading. (Byrne 2011: 207).

As a final point, it should be noted that Byrne acknowledges that the safety that the self-verifying nature of the doxastic schema brings about in self-ascribing beliefs is not sufficient to prove that reasoning according to the doxastic schema is knowledge-conducive. However, he emphasises that his proposal is still a plausible

way to form beliefs about our own beliefs. According to him, unless there is a compelling reason to believe otherwise, we should assume that this method of belief-formation is knowledge-conducive. In other words, until proven otherwise, we can rely on the doxastic schema as a way to gain self-knowledge of belief. (Byrne 2011: 207)

2. Intentional Actions

There are intentional actions and unintentional actions. Do we ever perform actions that are neither intentional nor unintentional?

(Mele 2012: 369)

Abstract

This chapter delves into the ongoing debate surrounding intentional action and proposes a novel theory: the Phronetic Theory of Action. This theory argues that an agent performs an action intentionally if she possesses the knowledge of how to perform the action and exercises that knowledge when performing it. While there are several other established theories of intentional action, including the Intentic Theory of Action (Simple View), Consequentialist Theory of Action, and Rational Theory of Action, this chapter demonstrates how the Phronetic Theory of Action provides a more comprehensive account of intentional action based on its focus on the role of exercising knowledge-how in doing something intentionally. In particular, this chapter discusses the challenges posed by different approaches to the Intentic Theory of Action and highlights how the Phronetic Theory of Action could offer a solution to those difficulties.

2.1. Introduction

The underlying purpose of this chapter is to provide a new account of intentional action. This account emphasizes the crucial role that the exercise of knowing-how plays in intentional action.

In the previous chapter, we explained the problem of self-knowledge and how the transparency account addresses this issue. We reviewed the literature on the transparency account in the work of Richard Moran, Mathew Boyle, and Alex Byrne. This chapter delves into another topic - intentional action - which raises questions about what it is and how it differs from unintentional action.

Having an account of intentional action is crucial in this research, as it allows for an explanation of knowledge of intentional action, which will be presented in the next chapter, and provides a means to address the problem of self-knowledge, particularly in the case of knowledge of one's own intention. Furthermore, in chapter four, it will be demonstrated how disregarding the true nature of intentional action, as presented in this chapter, and knowledge of one's own intentional action, as presented in the next chapter, renders Moran, Boyle, and Byrne's accounts of transparency implausible.

In the following sections, we will begin by briefly outlining our understanding of action, how we categorize intentional and unintentional actions, and introduce different theories of action. Next, we will present Anscombe's theory of action, which has shaped the current conception of intentional action. We will then introduce our own theory, the Phronetic Theory of Action (PTA), which highlights the critical role of knowledge-how and its application in identifying intentional actions. The rest of this chapter will cover the Simple View (Intentional theory of action) and some objections raised by cognitivists and consequentialists. Through our presentation of PTA, we will demonstrate that it not only offers a more comprehensive understanding of intentional action, but also provides proponents of the Simple View with the means to address these objections.

2.2. Intentional Actions

In our daily lives, we witness a variety of events, such as the rising and setting of the sun, the warmth it provides to the earth, the birth of new animals, the blooming of flowers, people falling in love, and students working on their thesis. These events can be categorised in various ways, and one common approach is to classify them as either actions or non-actions. Actions are generally considered to be events performed by agents, although it can be challenging to determine what or who should be classified as an agent. The concept of agency can be interpreted in different ways, leading to different entities being recognised as agents based on their unique properties.

To understand this complexity, attention to what happens during each sunrise and sunset and the interactions between the sun and the earth. The sun can be seen as the agent responsible for warming the earth. Here, the sun has an active role in warming the earth, and the earth has a passive role in receiving warmth from the sun. However, one can also question the passiveness attributed to the earth, as it could be viewed as an agent receiving warmth from the sun. Moreover, in an entirely different view, neither the sun nor the earth can take an agentive role, since none of them can decide on what they do. I do not want to elaborate on the different accounts of agency here. For my purpose, it is adequate to consider the less controversial level of agency in which only human beings who possess the required mental and physical abilities to bring about changes in the world would be counted as agents.

Although it seems that we have finally solved the difficulty of determining what we mean here by agent, questions remain regarding the changes an agent brings about. Can all an agent does be considered 'action'? Is an agent the one who brings about whatever changes in the world, or do there come only some particular changes into play when we are talking about an agent?

Changes brought about by human beings can be classified into different categories. One suggestion is to classify these changes into "act of human being" and "human act." In this classification, the former has a broader scope, including the latter¹³, although what makes human acts specific instances of an act of a human being is disputable. According to Anscombe, the main difference is that human actions are those acts of

¹³ See (Anscombe 1982: 13)

human beings that "are under the command of reason." What 'being under the command of reason' means is the issue we discuss later.

In addition, the changes caused by human beings can also be classified based on the different properties of the action or the agent. For example, voluntary and involuntary actions can be separated based on the characteristics of the agent during the action. Alternatively, basic and non-basic actions can be distinguished based on the properties of the action itself. The categorisation we will focus on here is the differentiation between intentional and unintentional actions. Some philosophers believe that such a distinction helps us define the concept of agency and discriminate between actions and non-actions. For example, in Davidson's view, intentional actions are the hallmark of agency. According to Davidson, an agent is someone who intentionally performs an action, or at least performs an action that can be described as intentional under some aspect. For him, actions are those events that, at least under one description, are done intentionally.¹⁴ (Davidson 2001: 46).

To explain what intentional actions are and what makes them separate from unintentional actions, I appeal to a way proposed by Carl Ginet (2004). In his paper '*Intentionally Doing and Intentionally Not Doing*,' Carl Ginet tries to define intentional actions by means of some traits of an action. If being an intentional action depends on a distinctive feature that only particular actions possess, then he holds that when one is doing something intentionally, the following general condition applies:

"S intentionally V-ed at T if and only if S V-ed at T and X." (Ginet 2004: 96).

Here, X represents a trait that makes S's action at T intentional.

By using Ginet's biconditional statement, we can categorise various action theories according to the definitions proposed by philosophers for intentional actions. These definitions can be sorted into three broad groups.

a. Intentional Theory of Action (strong): S intentionally V-ed at T if and only if S V-ed at T and S intended to V.

a. Intentional Theory of Action (weak): S intentionally V-ed at T if and only if S V-ed at T and S either intended to V or (at least) intended to try to V.

b. Consequentialist Theory of Action: S intentionally V-ed at T if and only if S V-ed at T and V was foreseen as certain or probable consequences of S's performance.

¹⁴ Anscombe is not happy with this idea that we can "introduce a restricted sense of 'action',... by trying to find a characterisation of a sub-class of events." (Anscombe 1982: 12)
See (Aucouturier 2021: 334)

b. Moral Theory of Action: S intentionally V-ed at T if and only if S V-ed at T and S was morally responsible for V-ing.

c. Rationalist Theory of Action: S intentionally V-ed at T if and only if S V-ed at T and S V-ed for a reason. (Anscombian)

There are two points to make here. First, there are other theories that we did not mention here. This may be because they are not as significant in the literature as the three theories we have just mentioned, or because they are based on these three theories. Second, despite the fact that these theories define intentional action in a completely different way, these theories are not necessarily in contradiction with one another, and there may be areas of overlap.

A point that has been overlooked in all of these theories is the importance of knowledge-how in the carrying out of intentional actions. To draw attention to this importance, in this chapter I add a new theory to this series: the Phronetic Theory of Action (PTA). According to this theory, S intentionally ϕ -ed at T if and only if S ϕ -ed at T and S (to some degrees) knew how to ϕ and S exercised this knowledge-how.

Before presenting PTA, we explain Anscombe's accounts of intentional action, through the Rationalist Theory of Action, in the next section. There are three reasons why we start with Anscombe's view. First, the dominant account of action belongs to her, and as Davidson says about the importance of her work, "Anscombe's *Intention* is the most important treatment of action since Aristotle."¹⁵ It is, therefore, very difficult to talk about intention, intentional action, and one's knowledge of intention and intentional action, without referring to her view. Second, in this dissertation, we agree with her in general on many of her ideas about intentional action and one's knowledge of one's own intentional action. Third, PTA and the other contributions we will make to the literature, most notably in Chapters 3 and 5, are in part formed on the basis of Anscombe's ideas.

¹⁵ On the cover of the Harvard University Press edition of *Intention* (Cambridge, Mass., 2000).

2.3. Rationalist theory of Action

In this section, we will explain the Rational Theory of Action, which is adhered to by philosophers such as Richard Moran, Mathew Boyle, Carl Ginet, Kieran Setiya (Setiya 2007, 2008), and David Velleman (Velleman 1991, 2000). These scholars largely follow Anscombe's views on intention and intentional action and are indebted to her ideas in this regard. Therefore, in this section, the focus will primarily be on Anscombe's theory of intentional action.

In the first section of her book "Intention" (1957), Anscombe distinguishes three concepts that can be employed by the term 'intention.' (Moran and Stone 2011: 34). Firstly, intentions could be revealed by *an expression of intention/intention for the future* ('I will do something'). Secondly, they could also be worked out by an *intentional action/intention in action* ('I am doing something'). Lastly, intentions could be disclosed by the *intention with which an action is done* ('I am doing something in order to do something else'). Accordingly, one can reveal one's intention to ϕ by expressing, "I will ϕ ," "I'm ϕ -ing," or "I'm ϕ -ing in order to ψ ."

Despite the differences, there is an overlap between these three constructions. The first one is dedicated to actions that have not yet been done, but the agent intends to do it in the future, the second type is assigned to actions that the person is doing, and the third type is allocated to actions that are being performed for doing something else that has not been done yet.

According to Anscombe and other philosophers of action, such as Davidson, only one of these notions plays a crucial role in explaining intention in general so that it can cover other notions of intention. For Anscombe, 'intentional action' is the key notion of 'intention,' with which one is able to explain the two other notions. In what follows, we explain Anscombe's account of intentional action.

2.3.1. Anscombe's account of intentional action

According to Anscombe, an intentional action is an action "to which a certain sense of the question 'why?' is given application; the sense is of course that in which the

answer, if positive, gives a reason for acting." (Anscombe 1957: 9). Hence, one has ϕ -ed intentionally, if and only if, later on, one is rationally able to give a particular kind of answer to the question 'why did you ϕ ?/why were you ϕ -ing?'. The specific sense this 'why?' question takes in Anscombe's account is that the answer provides reasons for an action, reasons without which that action would be unintentional. Drawing on Anscombe's primary assertion regarding intentional actions and the biconditional statement from Ginet, we can establish the following theory:

Rationalist Theory of Action: S intentionally V-ed at T if and only if S V-ed at T and S V-ed for a reason.

It is essential to mention two points here:

First, we refer to this approach as the Rational Theory of Action, since one of its crucial points is that an agent needs to have a reason for her intentional action; a reason which makes what she did justified and rational¹⁶. However, the type of reasons considered here only includes a specific category of reasons. Whatever is given in response to "why did you do this?" does not necessarily give a positive answer to the "why?" question in the particular sense intended by Anscombe, and it cannot be considered a reason that reveals the rationality of the action.

To further clarify, let's consider the following scenario: John is a virologist studying COVID-19, and Sarah is suffering from terminal cancer. Sarah asks John to end her life by infecting her with the virus, believing it could lead to a cure for COVID-19. John complies and infects Sarah while she is in a coma, ultimately resulting in her death.

If you were to ask John why he killed/transmitted the virus to Sarah, he would likely respond that it was to find a cure for COVID-19. While some may view this action as irrational or immoral, our definition of rationality is different. Rationality, in this context, refers to an agent having a reason for choosing a particular action over others, regardless of whether the action itself is right or wrong. In other words, even if an agent makes a mistake in her choice of action, for example, she has chosen a means that never reaches the goal, her choice could still be considered rational, as long as she thinks that she can achieve that goal in this way.

However, not all answers to the "why?" question demonstrate this kind of rationality. In the next chapter, we will explore how Anscombe identifies these types of answers.

Second, Anscombe recognises cases where, when asked about the motivation for

¹⁶ It's completely different from the 'rational agency' that Bratman talks about it.

an action, one may answer something like, "I had no particular reason," without the action being unintentional. However, she argues that these "for no particular reason" answers do not undermine her account, as they are not consistently relevant to the specific question she is investigating. In her opinion, "... it must not be supposed that because they can occur that answer would everywhere be intelligible, or that it could be the only answer ever given." (Anscombe 1957: 34).

2.3.2. One impediment and PTA's solution

For the rational theory of action, one impediment that we should consider is the cases of causal waywardness in which one has a reason for doing something, intends to do it, and will do it but not intentionally. For instance, consider the case of John, a virologist working on COVID-19, who intends to transmit the virus to Sarah, who is suffering from fatal cancer and has requested him to infect her in the hope that it could help John find a cure for COVID-19. Unbeknownst to John, he has already passed on COVID-19 to Sarah on a previous visit. Sarah dies of the virus just one second after John transmits it again. This case illustrates how an agent can have a reason to act intentionally, but external factors can cause the action to be wayward and unintentional. John had a reason to kill Sarah through the Coronavirus transmission, and he intended to do so, and he did so. Nevertheless, this was done by him unintentionally.

We find the objection to Anscombe's theory regarding causal waywardness invalid. This is because although John had transmitted the disease to Sarah, it had not been transmitted for the reason he had.

But what does it mean to say that we did something on the basis of the reason we had? In the case of John, he had a reason to kill Sarah, and he killed Sarah. How can we say that his act was not based on his reason? Our answer to this question is that although John knew how to transmit the virus and caused Sarah to die, he did not supervise the operation in light of the reasons he had for carrying this operation out. We advocate that although John knew how to transmit the virus and caused Sarah's death, he did not supervise the operation in light of the reasons he had for carrying it out. In other words, his action did not result from the exercise of his knowledge-how, and for this reason, it was unintentional. To provide a more thorough explanation, we

will present the Phronetic Theory of Action (PTA) in the next section.

2.4. Phronetic Theory of Action

In this section, we present our theory of intentional action, which we call the Phronetic Theory of Action. What should be taken into consideration in an account of intentional actions, but has been often overlooked, is knowledge-how. Intuitively, when a person performs an intentional action, she knows how to do it. In other words, having pertinent knowledge-how is a necessary condition for performing an intentional action. For example, suppose a person who has no experience with computers tries to connect to the Internet using a computer. She randomly taps on the keyboard and suddenly realises that the computer is connected to the Internet. Intuitively, this person did not intentionally connect to the Internet because she did not possess the necessary knowledge-how to perform this action.

Nonetheless, it is not sufficient to merely possess knowledge-how for an action to be intentional. This time, consider a person who knows how to connect to the Internet. She is writing her dissertation and randomly press the keyboard buttons in search of a specific punctuation mark. Ding-Ding! An email arrives, and she realises that she is connected to the Internet. Here, although she knows how to turn on the Internet, she did not do so intentionally. This is because she did not apply her knowledge here, and the connection to the Internet was by chance. It is worth mentioning that this does not mean that exercising knowledge-how and intentional action are identical. Exercising knowledge-how-to- ϕ is a necessary condition for ϕ -ing intentionally. Nevertheless, it is possible for one to exercise the knowledge-how-to- ϕ whereas one fails to ϕ .

One objection that may arise is that sometimes people try to do something without knowing how to do it, and if they succeed, it cannot be said that it was unintentional. Instead, it may be considered intentional. To further illustrate this point, let us revisit the example of connecting to the Internet. Imagine someone who initially does not know how to connect to the Internet, but tries to do so in order to send a copy of her dissertation to her professor. After several attempts, she successfully connected to the Internet and sent her draft. In this scenario, it appears that she acted intentionally, despite not knowing how to connect to the Internet at first.

The main idea here is to differentiate between intending to do something and trying to do something. As Hampshire says: “‘He is trying to do so-and-so’ already states the agent’s intention, with an added implication that there is some difficulty

and a possibility of failure.”(Hampshire 1982: 107). When someone is trying to do something, it implies that there is some difficulty and a possibility of failure. This implies that trying to do something only makes sense when someone has imperfect knowledge of how to do it. If someone knows how to do something, she will do it immediately when she intends to do so—it is senseless to say that she is trying to do it. If someone knows how to connect to the Internet, she will do so immediately when she intends to do so¹⁷. However, consider someone who could easily turn Wi-Fi on in the previous version of Windows but is now struggling due to an updated operating system. She may need to try to connect to the Internet. In other words, the knowledge that she had of how to connect to the Internet was knowledge of how to turn Wi-Fi on in the previous version of her computer's operating system. As the operating system is changed, although she may know to some extent what to do, she may not yet fully know how they can turn the Wi-Fi on in the updated version. The same can be said of a person who has been a swimming champion but has recently lost an arm. Obviously, the swimmer knows how to swim, but with both arms. With one arm, the champion certainly will need to practice.

The question that arises here is: what do we mean when we say that someone who does not know how to do something is trying to do it? We do not seem to attribute trying to connect to the Internet to someone who is simply taking eggs out of the fridge in the kitchen, or to someone who goes toward the phone to order pizza. Nevertheless, suppose this person picks up a book on ‘an Introduction to Internet Connection’ or goes to the phone and calls her friend, who is a computer expert, for guidance. In that case, it seems appropriate to say she is trying to connect to the Internet.

So, it does not seem irrelevant to say that trying to do something often involves 1) doing other things that we know how to do and 2) we believe that doing them leads us to bring about the thing that we do not know how to do.

However, the aforementioned conditions alone are insufficient to determine if someone is trying to do something. For example, consider a person who lacks the knowledge of how to connect to the Internet. This person goes to the phone to call a friend who is a computer specialist. Although she thinks that her friend can guide her through the process, during the call, she merely chats about the latest news and does not mention the Internet connection at all. In this case, it would not be reasonable to say that the person is trying to connect to the Internet. Therefore, in addition to the

¹⁷ Here, failure is possible, but it is another story to which we will come back later.

previously mentioned conditions, we require another criterion: trying to do something is doing other things that 3) we do in order to do the thing that we do not know how to do.

It may be argued that, in some cases, trying to do something does not necessarily involve doing other actions besides what is being attempted. One example could be trying to jump from one side to the other. Consider someone who attempts to jump a distance of two and a half meters between two chairs, having already successfully jumped two meters before. In this case, it appears that the person is only jumping when she is trying to jump two and a half meters, without doing anything else. This case is so akin to Davidson's carbon copier example. In his article "Agency," first published in (Davidson), Davidson brings an example in which a man is writing and making ten carbon copies, doing both intentionally, and yet not knowing that he is making copies, but just that he is trying to make them. Davidson's carbon copier knows how to make one, two, or even nine carbon copies. However, he is not sure how to make ten copies. He intends to do it, and he succeeds. However, the ten carbon copies were not made suddenly, 'just like that', but must be done carefully. The carbon copier expected that if he did again whatever he had done for nine copies, maybe by a little bit more pen pressure on the papers, he would make this time ten copies. Thus, we cannot say that Davidson's carbon copier lacks knowledge of how to make ten copies altogether, although a person who successfully makes ten copies every day may be more skilled. In other words, knowledge-how can come in degrees.

Indeed, having partial knowledge-how or degrees of knowledge-how does not imply that if someone does not know how to do something, she knows how to do something else instead. The difference in degrees also does not imply that one knows better how to do something else compared to what one is trying to do. Rather, it reflects a person's level of confidence and competence in performing an action. It simply means that a person can either have all the knowledge and skills necessary to accomplish the action in question or be less confident in her ability to complete it.

Regarding this matter, an individual who has previously jumped a two-meter distance and now attempts to jump two and a half meters possesses some knowledge of how to do it. This individual has experience in jumping a two-meter distance and is aware that more focus and power are required to succeed in jumping two and a half meters. She exercises her knowledge-how by employing the same technique she used to jump two meters, but this time with greater force. She does all this in order to jump two and a half meters.

From what we have expressed thus far, it may be apparent to some extent what we mean by exercising knowledge-how. As Anscombe says, it “is nothing but the doing or supervising of the operations” one is doing or going to do. (Anscombe 1957: 88). However, her use of “supervising” seems to refer to the supervision of others’ actions, such as a teacher supervising her students or a doctor supervising her nurses.¹⁸ Nevertheless, it is still possible to argue that we also have such supervision over our own actions. This means that we manage our actions, as well as the situations we find ourselves in, in order to successfully carry out our actions in light of the reasons we have for intending to do them. For example, when one intends to ϕ for any reason, and ψ -ing is inconsistent with ϕ -ing, one omits to ψ , or if situation X is implausible for ϕ -ing, one tries to avoid X.

Up to this point, we have established that the exercise of knowledge-how is a necessary condition for doing something intentionally, and we have explained what we mean by this. On the basis of what we have said and the biconditional statement from Ginet, we can present the following account of intentional action, which we call the “Phronetic Theory of Action” (hereafter PTA):

Phronetic Theory of Action (PTA): S intentionally ϕ -ed at T if and only if S ϕ -ed at T and S (to some degrees) knew how to ϕ and S exercised her knowledge-how.

In the upcoming section, I will begin by presenting various scenarios that describe different stories related to the transmission of Coronavirus from one person to another. These cases could be considered controversial examples of intentional actions. Then, I will introduce the Simple View (which I call Intentic Theory of Action (ITA)), which states that to perform an intentional action, one must have the intention to do so. After explaining the challenges that ITA faces when dealing with some of the scenarios, I resort to PTA to defend ITA against some objections.

¹⁸ Based on a comment that I received from Adrian Haddock and Eylem Özaltun in the conference ‘Rational Capacities in Thought, Perception, and Action’, May and June 2022.

2.5. Intentic Theory of Action

Consider the following cases:

Case 1: John had been self-quarantining for several days after being infected with Coronavirus. Despite *being aware of* the high risk of transmitting the virus to others and its potential lethality, John decided to invite Sarah over for a drink. He neglected to take any precautionary measures and proceeded to shake her hand and hug her as usual. As a result, Sarah contracted the virus from John, which ultimately led to her death.

Case 2: John *has the intention* to kill Sarah and *decides* to do so without leaving any trace that shows her death as a murder. He carefully infects Sarah with the CoronaVirus, which ultimately leads to her death.

Case 3: John is infected with CoronaVirus. He knows that Sarah had previously tested positive for antibodies and believes that she is immune to the virus. Out of curiosity, he *tries* to do his best to transmit the CoronaVirus to Sarah. Sarah gets the disease from John and passes away; unfortunately, as it became clear later, the test result was not accurate.

Each of these scenarios presents a controversial situation in which it is uncertain whether the agent acted intentionally or not. Different approaches can be used to answer the question of whether John intentionally killed Sarah, and these approaches do not necessarily agree with each other. Here we begin with one of the most ordinary approaches toward identifying intentional action to answer this question: the “Simple View¹⁹” or, as we will call it here, the Intentic Theory of Action (ITA). According to proponents of this view, having a particular mental state is closely involved in intentionally performing an action. While some philosophers argue that the mental state involved in intentional action could be a combination of different mental states, such as belief²⁰ and desire, ITA advocates believe that a distinctive state of intention is the very mental state one must have when one is doing something intentionally. This does not mean that intentional action occurs without any related desires or beliefs, but the specific role of intentions in intentional action cannot be reduced to the roles of desires and beliefs. Furthermore, ITA does not support the notion that having the

¹⁹ The term Simple View was first coined by Michael Bratman to cite the view that if one is ϕ -ing intentionally, one has the intention to ϕ . See (Bratman 1984: 377), and (Bratman 1987: 112)

²⁰ See (Davidson 1963)

intention to perform one action is sufficient to intentionally perform another action that is foreseen as a consequence of the first action.²¹ Therefore, based on ITA, an action is intentional if and only if two conditions are met: (1) the agent performs the action with the intention of doing so, and (2) the object of the agent's intention when performing the action intentionally must be the action itself. We can formulate ITA as below:

— S intentionally ϕ -ed at T if and only if S ϕ -ed at T and S intended to ϕ .

According to ITA, in the first scenario, killing Sarah is not what John intends, even if it is a foreseeable consequence of neglecting the precautionary measures. In this regard, based on ITA, if we can say that John killed Sarah in the first case, he did it unintentionally, even if one argues that he intentionally neglected the preventive measures. In the second scenario, however, the situation is different; John has a plain intention to kill Sarah. Therefore, according to ITA, we can say that he killed her intentionally.

Some philosophers find ITA untenable. There have been raised different objections that undermine the view from different angles. In the following subsections, we will review the objections of two main groups of critics and will show how PTA can defend ITA against these objections.

2.5.1. Cognitivism about intention (doxastic account of action)

Some philosophers, such as Carl Ginet (Ginet), Robert Audi (Audi), and Gilbert Harman (Harman), oppose the proponents of ITA and challenge the idea that the presence of a state of intending is necessary for doing something intentionally. They argue that an action could be intentional even if the agent had no explicit intention to perform that action. They support their claim by appealing to a thesis known as “cognitivism about intention” (Paul 2009: 2) or “doxastic account of intention” (Langton 2004: 244), which holds that having the intention to ϕ involves believing that ‘I am ϕ -ing’ or ‘I will ϕ .’ (Audi 1986: 18)

²¹ For clarity: ITA does not support the idea that having an intention for ψ -ing is sufficient to intentionally ϕ when ϕ is a foreseen concomitant of ψ -ing.

The argument presented by cognitivists against ITA can be formulated as follows:

- a) Intending to do something involves believing that one will do it. (Based on the doxastic account of intention)
- b) There are cases where the chances of achieving a favourable outcome are either zero or too slim.
- c) Therefore, in such cases, the agent either does not believe that there is any chance of success or believes that the chances of success are too slim.
- d) If so, in these cases, the agent either does not have an intention for what she is going to do or “does not flatly intend” to do that. (Harman 1976: 433). (Based on a and c)
- e) If the agent goes through an action of this type and succeeds, she did it intentionally without intending to do it.

2.5.1.1. One scenario, two approaches (1)

The standard cases that cognitivists present to show that a person would do something intentionally without ever intending to do it are similar to Case 3, in which John knows that the chances of transmitting the coronavirus to Sarah are either zero or too slim, but out of curiosity, he tries to transmit it to her and succeeds.

According to the cognitivists, in Case 3, John intentionally transmitted the Coronavirus to Sarah without explicitly intending to do so. They argue that John's lack of belief in the possibility of transmission makes it impossible for him to have an intention to transmit the virus. Therefore, they hold that John's action is intentional but not because he had the intention to transmit the virus.

Based on what was mentioned in section 2.5., an ITA proponent would say that in Case 3, John did not kill Sarah intentionally since he did not intend to do so. However, with respect to the transmission of the CoronaVirus to Sarah, an ITA proponent could have a different idea and consider the action to be intentional. However, if John does not firmly believe that he will succeed in transmitting the virus, how can an ITA proponent still consider the action intentional?

In our opinion, when John tries to transmit the CoronaVirus to Sarah, the very act of trying shows that he intends to transmit the CoronaVirus to Sarah, regardless of his

belief about the chances of success. Furthermore, even if an agent recognises a high probability of failure, she may still perceive a chance of success in her intentional actions. To further explore the connection between intending to do something and trying to do it, as well as the role of a minimal belief about our success in what we intentionally do, we need to examine the cognitivists' perspective regarding case 3 first.

2.5.1.2. Two confusions in the cognitivists' approach and their possible response to clear up the confusions

There are at least two confusions in the cognitivists' approach to Case 3 that require explanation. *First*, if John has no explicit intention to transmit the CoronaVirus to Sarah, why should we consider his action intentional? In other words, if an agent does not believe an action will be successful and, therefore, has no intention of performing it, why should we classify the resulting action as intentional if it is successful? *Second*, why would John attempt something without any belief in the possibility of success? It seems counterintuitive for an agent to continue with an action that has no chance of succeeding. These are the questions that the cognitivists must address in their argument.

Regarding the first confusion, Garcia's avowal provides us with a clarifying clue:

“In these cases the agents try to do something and succeed and it seems to me wrongheaded to deny that one does intentionally what one successfully attempts to do.” (Garcia 1990: 194)

Generally speaking, when an agent shows a deliberate effort to bring about a particular outcome and succeeds in doing so, we attribute intentionality to her action. In the scenario presented in Case 3, John made a conscious effort to transmit the virus to Sarah, and he succeeded in doing so. This is the primary reason why his action is considered to be intentional. Most philosophers acknowledge that intentional actions require some level of effort and volition on the part of the agent, and that success in achieving the desired outcome is a key factor in determining intentionality. With this in mind, many have attempted to develop an account of intentional action that can accommodate these types of actions.

Regarding the second confusion that the cognitivist needs to address, which is why someone would do/try to do something without believing she will succeed, philosophers have presented various reasons. For example, in his article '*Intentionally Doing and Intentionally Not Doing*,' Carl Ginet provides different reasons for a scenario in which a person, while not believing that she can move a boulder, attempts to move it by pushing on it and eventually succeeds. Ginet suggests that the individual may have the following reasons for her efforts:

“(1) to find out whether [s]he can move that boulder by pushing on it, (2) to demonstrate to someone else that [s]he cannot move it by pushing on it, or (3) to comply with a request or order to try to move it by pushing on it.” (Ginet 2004: 98)

He goes on and says:

“Having any of these reasons for pushing on the boulder is compatible with lacking the belief that one has a chance of moving it by pushing on it (though having the first is, of course, not compatible with having the belief that one cannot move it by pushing on it). And having any of them entails that one tried to move it by pushing on it and, therefore, that, if one thereby succeeded in moving it, one intentionally moved it by pushing on it. But intending to move it by pushing on it is not compatible with lacking the belief that one has a chance of moving it by pushing on it.” (Ginet 2004: 98)

In this scenario, the agent attempts to move the boulder despite not believing that her attempt will succeed. According to the doxastic account of intention, lacking this belief is not compatible with having an intention to move the boulder. However, if the agent succeeds in moving the rock, Ginet considers the action to be intentional.

2.5.1.3. At least a partial belief in success is necessary to do something intentionally

It is important to reiterate that the reasons attributed to the agent by Ginet are motivations for *trying* to move the rock, rather than for simply moving it. Let us examine each of the reasons he presents for an agent who attempts to move the rock despite not believing it to be possible. We will start with the last two reasons before returning to the first one.

In his article, Ginet acknowledges that having the second or third reason for pushing on the boulder is compatible with the belief that one cannot move the rock

by pushing on it. Shortly after mentioning the possible reasons for trying to move the boulder, Ginet stipulates a condition for intentional actions. The interesting point is that he himself believes that this condition only applies when the person finds it, albeit very unlikely, still possible that she can move the boulder. He says:

Note that this sufficient condition is entailed by cases of S's intentionally V-ing by A-ing where S intended to V by A-ing but believed only that she might succeed in doing so. Suppose, for example, that when S pushed on the boulder she was far from confident that she could thereby move the boulder but she believed - correctly, as it turned out - that there was a chance that her strength was up to the task. (Ginet 2004: 100-01)

Thus, it should be noted that although Ginet asserts that having the second and third reasons is compatible with believing that moving the boulder is impossible, he also acknowledges that if the person succeeds in moving the boulder and has done it intentionally, indeed she had not found it impossible to move the boulder. Instead, she believed that it was challenging and perhaps beyond her capability. Here we can determine a necessary condition for doing something intentionally, which we call the doxastic requirement:

The doxastic requirement: if S ϕ -ed intentionally, S believed to some extent that there was still a possibility, albeit a small one, for ϕ -ing.

It states that if someone intentionally performs an action, she must have believed, to some extent, that there was still a possibility of achieving that action, even if the likelihood of success was small. This requirement ensures that intentional actions are not simply a matter of blind luck or chance but are grounded in a rational belief that the action is possible to achieve, even if success is uncertain. In the following subsection, we will explain more about the rationale behind the doxastic requirement, and we will see how PTA supports it.

2.5.1.4. How PTA addresses the cognitivists' cases

Regarding the second and third reasons for moving the boulder, as presented by Ginet, if the person who is pushing the boulder thinks there is no chance of moving it, it is odd even to say that she is trying to move the boulder. In such a case, it is hard to say that she intended to stage a demonstration or comply with a request to move the

rock, as she does not believe that either of these actions would cause the rock to move. Therefore, moving the rock is not even a foreseen concomitant of what she is doing²². Here, it is more plausible to say what she is doing is demonstrating to someone else that she cannot move the boulder by pushing on it, or she is complying with a request or order to push the rock. If the rock is moved in this scenario, it is an unexpected result beyond her will, and she would likely be surprised by her success. Therefore, when she moved the rock, she must have been shocked. In such circumstances, most philosophers of action concede that it is implausible to say that one intentionally did the action when she succeeded unexpectedly.²³ It supports the doxastic requirement, according to which an agent must believe that there is still a possibility, albeit small, of accomplishing the action she is attempting to do intentionally.

PTA can help us clarify the situation here. From the standpoint of PTA, in such a case, even if we assume that she knew to some extent how to move a boulder (e.g., by having moved smaller boulders before), it cannot be said that she exercised this knowledge in this case. Instead, she exercised the knowledge of how to demonstrate to someone else that she cannot move the boulder by pushing on it or how to comply with a request or order to push a rock. This is because she pushed the boulder not to move it but to show her friend that she could not move it or to comply with the request. She supervises certain operations when she shows someone she cannot move the boulder. Although the operations she supervised are similar to those she would have supervised if she were intentionally moving the boulder, this does not necessarily mean that she intentionally moved the boulder in the actual scenario. (she did not push the rock in order to move it, but she pushed the rock in order to do something else). Therefore, according to PTA, she did not move the boulder intentionally.

Now let us back to the first reason attributed to the agent by Ginet for trying to move the boulder: she is pushing on the boulder to find out whether she can move it. In this case, we can sympathise with Ginet and say that the agent moved the boulder intentionally. Unlike the previous two cases, this description that the agent is curious to find out whether she can move the boulder is incompatible with her believing that she cannot move it and is compatible with her seeing a slim chance of moving it.²⁴ In

²² If moving the boulder was a foreseen concomitant of what she was doing, then the story would be different, which we will explain in the next part, which is about the double effect doctrine.

²³ Based on Anscombe's view, followed by most rationalists, when one becomes shocked by what one did, it entails that one's knowledge of one's own action must be observational, which means one did it unintentionally. We will explain it in more detail in the next subsection.

²⁴ What about a suspension of belief? When we are trying, and we are doing ψ in order to ϕ ,

this situation, she would find it rational to give it a try and do her best to move the rock. In other words, she has a partial belief or minimal confidence that she can move the rock by trying to push it, and this confidence justifies her attempt to move the boulder.

From the standpoint of PTA, we can say that she exercises her knowledge-how-to-move-the-boulder by pushing the boulder as hard as possible in order to move it. The same could be true about Case 3. Although John is not sure whether he can transmit the virus to Sarah, we can say that he does not believe that it is totally impossible to make her sick, and his little confidence justifies his attempt to transmit the virus to Sarah. If we assume that John does not have this minimum confidence, then it is hard to call his action intentional.

2.5.1.5. The cognitivist's attempt to reject ITA fails

Based on the descriptions provided, it appears that the cognitivist's attempt to reject ITA has failed, and Garcia's statement that "what I try to do I intend to do" holds true. (Garcia 1990: 199). To explain this point, he says:

"When I try to do A (e.g., try to phone Jones), I do some B (e.g., push buttons on a telephone) with the intention that doing B constitute, or constitute a part of, my doing A. To say my doing B was an attempt to do A is to say what I intended my doing B to be (or be a part of) and therefore to say something of why, with what aim, I did it. Thus, we often say what we were trying to do as a way of stating the aims of what we did." (Garcia 1990: 199)

Garcia argues that when we try to do something, we do certain actions with the intention that they will be part of achieving our ultimate goal, and this is why we often state what we were trying to do as a way of stating our intention. If John or the agent attempting to move the boulder are genuinely making an effort to accomplish something, they have the intention to do the thing. Furthermore, according to the explanation we gave earlier about trying, they believe to some extent that it is possible to do the thing. Accordingly, given this extent of confidence in the agent who is trying to do something, it can be said that *they also have the intention* to accomplish it, without

although we can have a suspension of belief regarding our abilities, we cannot have a suspension of belief regarding what we are doing.

violating the "doxastic account of intention."

A tenacious cognitivist may argue that, based on the doxastic account of intention, having an intention to do something (ϕ) necessitates believing that "I am ϕ -ing" or "I will ϕ ." Consequently, having only partial confidence in ϕ -ing, as in Case 3, would not suffice to intend to ϕ . To address this objection, we must examine the difference between believing that "I am ϕ -ing" or "I will ϕ " and having a partial belief that "I would ϕ ." Is the only distinction that, in the latter, one is more likely to fail in what one intends to do? If the problem of potential failure is an obstacle to intending, it can undermine the doxastic account of intention entirely. After all, when an agent believes that "she will ϕ ," there remains the possibility of not succeeding in ϕ -ing.

It is important to note that some proponents of ITA have responded to cognitivist criticisms by offering a weaker version of their original theory of action, which states that:

— S intentionally ϕ -ed at T if and only if S ϕ -ed at T and S either intended to ϕ or (at least) intended to try to ϕ .

The rationale for presenting this weaker theory is that if the cognitivist's claim that an agent may act intentionally without intending to do so was valid, it could be argued that if the agent attempted to do so, she intended to try to do so. However, if we accept our earlier definition of trying to do something as doing ψ in order to ϕ , then it appears that we no longer require this weaker version of ITA. Attempting to do ψ in order to ϕ but lacking the intention to ϕ appears absurd.

2.5.2. Consequentialist and Moral Theories of Action

Another objection to ITA is presented by those I refer to as consequentialists. This group of critics argues that ITA improperly limits the object of intention solely to the specific action ϕ when performing ϕ intentionally. They believe that the action ϕ can still be considered intentional even when it is simply an unintended byproduct of ψ , an action that one intended to perform. In such scenarios, a person would carry out ϕ intentionally without explicitly intending to do so, if ϕ occurs as a result of ψ and the individual intended to perform ψ . The main justification for labelling ϕ as an intentional action is the foreknowledge that ϕ is a certain or probable outcome of ψ . According to consequentialists, actions executed with this knowledge are intentional.

However, we will contend that this view is flawed, as the unintended consequences of an intended action should be categorised as either unintentional or non-intentional.

2.5.2.1. One scenario, two approaches (2)

The standard cases that consequentialists present against ITA are similar to Case 1, where John does not intend to endanger Sarah's life and merely wants to have a drink with her. However, Sarah's death could be seen as an unintended consequence of John's interaction with her, given that he is aware of the high risk of her contracting the virus.

According to consequentialists, in Case 1, even though the transmission of the virus was an unintended concomitant of inviting Sarah, John still infected her intentionally. They argue that John's actions led to Sarah's death intentionally because he was aware of the high risk of infection and the possibility of her dying from it.

In contrast, as mentioned in Section 2.5., an ITA advocate would assert that, in Case 1, John did not kill Sarah intentionally, since he had no intention of doing so.

2.5.2.2. Two confusions in the consequentialists' approach and their possible response to clear up the confusions

There are at least two points of confusion in the consequentialists' approach concerning Case 1 that need clarification:

First, it is essential to understand how we, in general, and consequentialists, in particular, differentiate between intended and unintended events within a set of occurrences. We need a clear method for categorising a series of events as either intended actions or their foreseen/unforeseen consequences.

Second, if John has no intention of transmitting the virus to Sarah, why should we regard his action as intentional? We mentioned that it could be because John is aware

of the high risk of infection and the possibility that she might die from it. However, the question remains: why should awareness play a role in determining intentionality in such cases?

Consequentialists often overlook the first confusion and merely assume that John intended one outcome while having no intention for another. In their examples, if they attempt to address this issue, they typically claim that 'the person herself states that she did not have such an intention', or they provide a similar explanation. However, in our opinion, this is a crucial question: why do we say that John intended to drink with Sarah, but had no intention of making her sick or causing her death?

To further illustrate the first confusion, consider a marathon runner²⁵ who knows that if she participates in a competition, her sneakers will be worn out along the course. What she intends to do is participate in the competition and run all the way to the finish line. Wearing out her shoes or wetting them when she accidentally steps into a puddle are not actions she intends to perform. Both are consequences of her intended action—she can foresee the former but not the latter. The question now is, why do we say that these are merely consequences of her intended action and not actions she intends to perform? Before answering this question, let's discuss a little more the second confusion.

Regarding the second confusion, i.e., why we should consider an unintended concomitant of an intended action to be intentional, consequentialists can present at least two theories of action:

Consequentialist Theory of Action: S intentionally V-ed at T if and only if S V-ed at T and V was foreseen as certain or probable consequences of S's performance.

Moral Theory of Action: S intentionally V-ed at T if and only if S V-ed at T and S was morally responsible for V-ing.

According to the Consequentialist Theory, in Case 1, John made Sarah sick intentionally, since it was a foreseen consequence of John's intended action. Similarly, as the moralist suggests, John made Sarah sick intentionally, since he is morally responsible for her illness.

A famous example used to advance this line of argument is the Tactical Bomber case²⁶. A bomber pilot intends to bomb an enemy munitions factory, which happens to be located next to a school. The pilot knows that by bombing the factory, he will

²⁵ See (Bratman 1984: 123)

²⁶ See (Bennett 1981: 96)

also destroy the school and kill the children inside the building. He is distraught about the loss of innocent lives and, on the surface, does not have such an intention to kill the children. Nonetheless, he proceeds to drop the bombs on the factory. In this case, although he did not intend to kill the children, some philosophers maintain that he intentionally caused the deaths of the students.

In opposition to this line of argument, proponents of ITA claim that the term 'intentional' is not being applied accurately. They argue that when an agent knowingly causes harm, she is responsible for her wrongdoing. By saying that "she intentionally causes the harm," we mean that she is blameworthy²⁷.

However, opponents of ITA, like Carl Ginet, still take issue with this line of defence. According to Ginet, this argument lacks force, since "S [the agent] knowingly caused the harm" is sufficiently censorious that we do not need necessarily to assert that S did it intentionally. (Ginet 2004: 97).

In this discussion, both the Consequentialist and the Moral theories emphasise the point that Ginet mentioned: "S [the agent] knowingly caused the harm." 'Knowingly doing something' is the cornerstone of these theories. Here, it can be argued that the Moral Theory can be reduced to the Consequentialist theory because it can be interpreted as follows:

- 1) Whatever we are morally responsible for doing, we did knowingly,
 - 2) whatever we did knowingly, we did intentionally (Consequentialist Theory of Action)
- ; so
- 3) Whatever we are morally responsible for doing, we did intentionally.

2.5.2.3. How PTA addresses the consequentialists' cases

Appealing to our idea about exercising knowledge-how (PTA) may help address both

²⁷ Such an answer, in fact, undermines the Moral Theory of Action. (I do not intend to go into this theory here, perhaps in the second version of Chapter Two I will expand on this theory and also address the the Doctrine of Double Effect). In general, the idea that whenever we are morally responsible for something, then we did it intentionally does not seem right. But the opposite is more acceptable and reasonable. One is morally responsible for what one did intentionally. A reason that can be given for this, is the active role of the agent in doing what happened intentionally. On the other hand, the assumption that we can be morally responsible for something without doing it intentionally depends to a large extent on what moral law we follow.

confusions. Regarding the first confusion, the runner tries her best to cross the finish line first. She strives to take longer strides and maintain a straight path as much as possible. However, this is not the case with wearing out her shoes or getting them wet. If she realizes she needs to increase her speed to win the race, she will certainly do so, but if she notices that her shoes are more durable than she initially thought, she will not make an effort to wear them out. The same applies to wetting her shoes; if she sees that her shoes are not wet, she will not actively seek a puddle to step in. In short, she exercises her knowledge of how to run and win a race but not how to wear out or soak her shoes. She supervises her actions to achieve the goal of crossing the finish line first, but she does not supervise them to wear out or wet her shoes. In general, it can be said that **intention is accompanied by a commitment that can manifest itself in exercising knowledge-how.**

Regarding the second confusion, the mere fact that the pilot knows that bombing the factory would result in the deaths of the children does not seem to be a convincing reason to claim that the pilot intentionally killed the students. An agent may be aware that something is very likely to occur as a consequence of her intended action; however, what the agent intends to do is not meant to be carried out *in order to* do the other event. In other words, in such cases, knowledge of how to perform the concomitant action is not exercised.

The consequentialist might argue that this is why it is said that the agent did not intend to do this, but it is not enough to assert that the agent did not do it intentionally. Nevertheless, it seems that it is consequentialist that needs to provide a reason for considering the foreseen consequences of intended actions to be intentional. As Mele suggests, there is no need to label actions only as intentional and unintentional; rather, they can also be classified as non-intentional:

“[I]nsofar as an agent who is A-ing is neither aiming at A-ing nor trying to A, either as an end or as a means to (or constituent of) an end, she is not intentionally A-ing; insofar as an agent is A-ing knowingly and non-accidentally, she is not unintentionally A-ing; and actions that are neither intentional nor unintentional are non-intentional.” (Mele and Sverdlik 1996: 274)

According to Mele, actions that do not fit the criteria for being intentional or unintentional are considered non-intentional. Therefore, as long as the consequentialist has not provided a reason for her claim, it can be said that the unintended consequences of an intended action are either unintentional or non-intentional.

3. One's Knowledge of One's own Intentional Action

A man has practical knowledge who knows how to do things.

(Anscombe 1957: 88)

Abstract

In this chapter, we discuss one's knowledge of one's own intentional action. According to Anscombe, when one is doing something intentionally, one is doing it knowingly. Furthermore, she argues that our knowledge of our own intentional actions is non-observational. Despite Anscombe's insights, she does not provide a clear explanation of how non-observational knowledge is possible. This puzzle has led to numerous debates and inquiries into the nature of such knowledge. After explaining Anscombe's ideas, we will show how Kieran Setiya defends the possibility of such knowledge and how Setiya's account is consistent with PTA.

3.1. Introduction

The underlying purpose of this chapter is to explore the nature of one's knowledge regarding one's own intentional actions. Anscombe argues that this knowledge is non-observational and practical, which distinguishes it from other types of knowledge and makes it partially similar to self-knowledge. As Thompson mentioned, "the overarching thesis of *Intention*²⁸ was that self-knowledge in this familiar sense extends beyond the inner recesses of the mind, beyond the narrowly psychical, and into the things that I am doing". (Thompson 2011: 200). It could be said that Anscombe expanded the concept of self-knowledge to encompass not only the inner workings of the mind but also external actions. Although the close resemblance between knowledge of one's own intentional action and self-knowledge²⁹ is worth examining, what makes the analysis of the former knowledge more important in our research is the role it can play in addressing the problem of self-knowledge, particularly in regard to knowledge of one's own intentions.

In the next chapter, I will demonstrate how overlooking the true nature of intentional action, as presented in the previous chapter, and the knowledge of one's own intentional action, as introduced in this chapter, undermines the plausibility of Moran, Boyle, and Byrne's accounts of transparency. Furthermore, in chapter five, I will explain how this knowledge can help one form justified beliefs about one's own intentions.

3.1.1. Why are we beginning with Anscombe? What is

²⁸ Anscombe's well-known book.

²⁹ It seems that the asymmetry we discussed in the first chapter can also be adopted for the knowledge that one has of one's intentional action. To better understand this asymmetry, consider the following example: If Sarah is writing something, John knows that through some behavioral evidence, like observing that Sarah moves her hand on a paper when she has a pen in her hand or through Sarah or somebody else's testimony about what she is doing now. Although John needs the very same evidence to know about his own action when he is doing it unintentionally, it seems that if John is writing something intentionally, he knows that without resorting to such evidence. Moreover, John could be wrong about the action that is ascribing to Sarah. Maybe Sarah is merely drawing something when she is moving the pen on the paper. However, it is seldom the case that John is wrong about identifying the actions that he attributes to himself.

the starting point?

In this research, the knowledge one has of one's own intentional action plays a pivotal role in explaining how one has privileged access to one's own intentions. As we mentioned earlier, in this dissertation, we concur with Anscombe on many of her ideas about intentional action. Regarding the knowledge of one's own intentional action, we also follow her account. In this regard, it is necessary to explore Anscombe's account of the knowledge of one's own intentional action to understand the characteristics of this knowledge. Additionally, we believe that PTA can lend support to defending Anscombe's account of knowledge of one's own intentional action against challenges that have been raised about it.

In the previous chapter, we discussed Anscombe's view that an intentional action is an action "to which a certain sense of the question 'why?' is given application; the sense is of course that in which the answer, if positive, gives a reason for acting." (Anscombe 1957: 9). To clarify when the "why?" question in the required sense has application and what doing an action for a reason means, Anscombe attempts to specify the answers to which the certain sense of the question "Why?" does not apply. According to her, the question "Why?" does not have an application in the following cases:

1. "I was not aware I was doing that'." (Anscombe 1957: 11). According to Anscombe, when one is doing something intentionally, one is doing it knowingly. For example, if someone is sawing a plank and is asked, 'Why are you sawing Mr. John's plank?' and she responds, 'Sorry! I didn't know it's Mr. John's plank!', then the agent's act of sawing Mr. John's plank was unintentional.

2. "I knew I was doing that, but only because I observed it'." (Anscombe 1957: 14). According to Anscombe, one's knowledge of one's intentional action must be of a "non-observational" type. If someone is sawing a plank and is asked, 'Why are you making a squeaky noise?', and she becomes aware of the noise and say, 'Yes! It's like I'm making a lot of noise!', then the agent's act of making noise was unintentional.

3. "It was involuntary!" (Anscombe 1957: 12). Sometimes, we may even have non-observational knowledge of our action, but perform the action involuntarily. Suppose someone is sawing a plank. When she pulls her arm back, her hand hits a nail, and she sighs in pain. Someone asks, 'Why are you shouting?' She answers, 'It

was involuntary!' Here, she has non-observational knowledge of her action. However, she knows the cause of her action only through observation (if she has realized that hitting a nail causes the pain). (Anscombe 1957: 15). Such an involuntary action is unintentional in Anscombe's view.

4. "I saw such-and-such and it made me jump!" (Anscombe 1957: 16). Sometimes we may give an answer to the question 'why?.' But this answer still cannot provide a reasonable reason for doing the relevant action. Suppose someone is sawing a plank. Meanwhile, a cat passes in front of her and frightens her. She thinks that the thing that passed in front of her is a demon. Someone asks her, 'Why did you jump back at once?' She answers, 'I saw a demon!' Here, the object of her fear is the cat, but the reason for her fear is that she thinks that creature is a demon. She says that she has seen a demon, but she certainly has not seen that the 'demon' is the cause of her fear. The main reason for her panic is that she considered what she saw to be a demon. According to Anscombe, if things such as thought, desire, feeling, seeing, hearing, etc. cause an action, they are 'mental cause'. According to her, if mental causes explain why someone did something, that action is unintentional.

Among these negative answers, the most important aspect for this study is the requirement of knowing one's own action when performing it intentionally. Examining the negative answers reveals a critical point in the first answer: when we act intentionally, we do so knowingly. It is impossible to intentionally interrupt someone without knowing that you are interrupting them. As Hampshire puts it: "doing something with intention, or intentionally, entails knowing what one is doing." (Hampshire 1982: 102). Another important point is the characteristic that Anscombe attributes to one's knowledge of intentional actions in the second type of answers mentioned above: this knowledge is non-observational. According to some philosophers, such as David Velleman, Kieran Setiya, and Sarah Paul, the non-observational nature of one's knowledge of intentional action is the most challenging characteristic of this knowledge.

3.1.2. What is the challenge and how do we address it?

As previously mentioned, according to Anscombe, when one is performing an action intentionally, not only is it done knowingly, but it is also done non-observationally.

This idea quickly became controversial. Some philosophers questioned the necessity of such knowledge for intentional action, while others challenged the non-observational nature of this knowledge, arguing that it is impossible to have such non-observational knowledge of action that is not based on one's perceptions.

We concur with Anscombe that when one is performing an action intentionally, one is doing so knowingly and with non-observational knowledge. In defence of the possibility of such non-observational knowledge, we follow Kieran Setiya, who proposes that to intend an action is, at least in part, to believe that one is doing it or that one will do it, a belief justified by one's knowledge of how to perform the action. (Setiya 2008).

In this chapter, I will begin by briefly reviewing Anscombe's account of knowledge in intentional action. Next, I will discuss the two-factor view and Anscombe's understanding of practical knowledge. Finally, I will present Setiya's analysis of non-observational knowledge of action and its entanglement with PTA.

3.2. What one is doing intentionally, one is doing knowingly!

As mentioned earlier, according to Anscombe, a typical answer that definitively denies any intention in doing something is, 'I was not aware I was doing it.' By giving such an answer, the 'why?' question in the required sense is claimed to have no application (Anscombe 1957: 11). For instance, if a colleague is not aware that she is distracting you with the click-clack of typing, she cannot be intentionally interrupting your attention. On the other hand, unless she is in some altered state of consciousness, we expect her to be aware of her typing. There are other situations in which one can be unaware of what one is doing. Your colleague may be intending to interrupt your attention without knowing that she is producing a famous African rhythm with the click-clacks of typing, simply because she does not recognise this melody. About this kind of unawareness, Anscombe says:

"Since a single action can have many different descriptions, ..., it is important to notice that a man may know that he is doing a thing under one description, and not under another. ... So to say that a man knows he is doing X is to give a description of what he is doing under which he knows it." (Anscombe 1957: 11-12).

An action definitely could be subjected to various descriptions, and according to Anscombe, the "why?" question can have an application only under some of these descriptions. More generally, in her opinion, we cannot provide reasons for our actions if we do not even know that we are doing those actions.

3.2.1 This knowledge is non-observational

Whenever we recognise that something is happening, we provide a description of it. Take, for example, the scenario presented above. Numerous individuals in the room are aware of what is going on: those who hear the clicking know someone is tapping the keys of the keyboard harshly; some of them know who is doing that because they see her. If they also notice you, they probably understand that your attention is interrupted each time you turn your head or simply react reflexively. If they, too, are

bothered by the sound, they are undoubtedly aware of their annoyance, just as you are aware of your discomfort. Finally, they might know both of you well enough to judge whether her intention is to distract you or merely to finish her work quickly. Given these various forms of knowledge, can we differentiate our knowledge of our own intentional actions, which permits us to provide a relevant response to Anscombe's "why" question, from all other types of knowledge? According to Anscombe, our knowledge of our intentional actions is not the same as other types of knowledge that involve observation:

"... in so far as one is observing, inferring, etc. that Z is actually taking place, one's knowledge is not the knowledge that a man has of his intentional actions. By the knowledge that a man has of his intentional actions, I mean the knowledge that one denies having if when asked, e.g., 'Why are you ringing that bell?' one replies 'Good heavens! I didn't know I was ringing it!'" (Anscombe 1957: 50-51)

It seems evident that your colleague is acutely aware of her typing. As Anscombe puts it, she does not think, 'Let me see, what my finger movements are bringing about? Ah yes! typing.' (Anscombe 1957: 51). In contrast, the knowledge you and other colleagues in the office possess about the situation primarily relies on what you perceive and other evidence upon which you make inferences.

What about your knowledge about your moving by reflex each time your colleague strikes the keys? You know quite rightly what you are doing without resting on any observation. Should we conclude that your behaviour constitutes an instance of intentional action? According to Anscombe, this conclusion is erroneous since you would likely assert that it was involuntary if asked later on, thereby rejecting, as we mentioned earlier, the application of Anscombe's "why" question (Anscombe 1957: 12). To exclude those types of actions that we do involuntarily or by reflex, Anscombe provides further elaboration on what she does not consider as non-observational knowledge: "Now among things known without observation must be included the cause of some movements. ... But in examples [like moving by reflex] the cause of motion is known only through observation." (Anscombe 1957: 15).

Consider again your involuntary movement caused by the rattles. Your knowledge of its cause is based on your simultaneous perception of the noise produced. You are moving unintentionally. Compare this with the case you turn your head voluntarily toward the noise source at the person who produces it. This time, needing no observation, you know why you are moving intentionally. As for the noisy colleague, utterly unaware of you and your embarrassment, she continues beating the

keys, and it is without observation that she knows why she is doing that.

Thus far, we have established that, according to Anscombe, our knowledge of our own intentional actions is non-observational³⁰, as opposed to our knowledge of unintentional actions or the actions of others. In the next step, we will explore the possibility of non-observational knowledge in more detail. However, before doing so, it is crucial to address the challenges associated with this type of knowledge. To elaborate on these challenges, we will reference a theory that partially agrees with Anscombe's concept of non-observational knowledge. This theory suggests that non-observational knowledge pertains to one's knowledge of one's will/intention, rather than one's knowledge of what happens in the outside world resulting from one's actions.

3.2.2. It is hard to accept that the knowledge is non-observational

It is easier to accept that our knowledge of our bodily movements is acquired without observation. For instance, to perform basic actions like raising a hand, clenching a fist, or bending a knee, it is unnecessary to observe and witness what is going on in the world to ensure successful execution. However, in her book, Anscombe provides examples of everyday actions, such as opening a window, writing on a blackboard, and so on, and argues that when we do these actions, our knowledge is not acquired through observation.

The possibility of knowing these types of actions in a non-observational manner is debatable because such actions not only alter the position of our body parts but also bring about changes to the external world. (Schwenkler 2011: 139). Anscombe argues that one who is writing something knows what one is writing, even with one's eyes closed, without looking at one's writing. However, it is evident that bending the knees

³⁰ Here we should consider Setiya's remark that "Anscombe does not deny that knowledge of what one is doing intentionally typically depends on empirical or other knowledge of the world. In the example of paralysis, I cannot know that I am clenching my fist in doing so intentionally unless I know that I have recovered. But even when I have that knowledge, I do not know that I am clenching my fist on the basis of sufficient prior evidence. Likewise, in the well-known vignette from Intention, one cannot know that one is pumping water into the house unless one knows that the equipment is working properly; but when all goes well, one's knowledge of what one is doing is not perceptual or inferential." (Setiya 2011: 173)

and writing with closed eyes are two vastly different kinds of actions in terms of the changes they cause. Anscombe acknowledges this point and notes that the following objection will likely come to the reader's mind:

'Known without observation' may very well be a justifiable formula for knowledge of the position and movements of one's limbs, but you have spoken of all intentional action as falling under this concept. Now it may be e.g. that one paints a wall yellow, meaning to do so. But is it reasonable to say that one 'knows without observation' that one is painting a wall yellow? And similarly for all sorts of actions: any actions that is, that are described under any aspect beyond that of bodily movements. (Anscombe 1957: 50)

In short, as Moran mentions, we can say that it is reasonable to think that it is not possible to be aware of making such changes in the external world without observation. (Moran 2001: 46). In the following section, we will present a possible solution proposed by Donnellan to address this issue. Donnellan's viewpoint is relevant to this research as it provides a better understanding of the Problem of Failure, which targets non-observational knowledge of action. Additionally, Donnellan's perspective serves as a useful transition to Anscombe's idea concerning the difference between theoretical and practical knowledge.

3.3. Two-Factor View

Some philosophers divide intentional action into two components: "interior" and "exterior." The interior part includes the agent's relevant will or intention, while the exterior part consists of the consequences of this will and intention in the external world. (Bratman 1984), (Searle 1980), (Davidson 1971). Based on this perspective, some philosophers argue that the agent possesses knowledge of the interior part of the intentional action without observation, while knowledge of the exterior part requires observation. This view, known as the two-factor view, is challenged by Anscombe. In what follows, we will discuss the two-factor view and Anscombe's critique of it.

Keith Donnellan (Donnellan) makes a distinction between one's knowledge of one's own intention and one's knowledge of the results of one's actions. Donnellan argues that knowledge of one's own intention is necessarily immediate and non-observational, while knowledge of the results of one's actions could be mediate and observational, relying on one's perception of what one's action brings about. He suggests that Anscombe's claim about non-observational knowledge of intentional action primarily pertains to knowledge of one's intention, and knowledge of the results of an action may be obtained through observation.

However, Anscombe criticises this distinction between knowledge of one's intention and knowledge of one's action and describes it as "a mad account". (Anscombe 1957: 52). She argues that non-observational knowledge of one's own intentional action includes not only knowledge of one's intention but also knowledge of the results of one's action. According to her, this distinction reduces the role of the agent to a mere subject who only has the power to intend, while the results are outside the agent's control and merely the "grace of fate" (Anscombe 1957: 52). Anscombe emphasises that when one performs an intentional action, one does not intend to do it and then wait to see what happens. Rather, one intends to raise one's hand, and one knows with certainty that one's hand will be raised in accordance with one's intention.

It should be noted that our thesis on exercising knowledge-how, i.e., PTA, also supports Anscombe's view. When one intends to perform an action, one is utilising one's knowledge of how to carry out the action, which involves overseeing and directing one's operations to achieve the intended outcome. Consequently, we can argue that the outcome, which is the object of the agent's knowledge, is a result of

what the agent did/is doing, rather than a matter of chance. This feature is what Anscombe refers to as *practical knowledge* and distinguishes it from *theoretical knowledge*, a topic we will delve into further in the following section.

3.4. One's Knowledge of One's own Intentional Action is Practical Knowledge

Donnellan's remark on one's knowledge of one's intentional action pertains more to complex actions in which the probability of failure is high. The possibility of error or failure in such complex cases is the reason for considering a distinction between these two types of knowledge. Although we may not agree with Donnellan's view on the difference between obtaining knowledge of intention and knowledge of intentional action, we can acknowledge that when we fail, the actual outcome of our intention may deviate from our intended outcome.

What if we notice our mistake while performing the action? Does failure to fulfil our intention mean that we do not have non-observational knowledge of our actions? Does the problem of failure not indicate that we only know the success of our actions through observation? Anscombe is aware of this problem of failure and acknowledges that we may not always succeed in completing our intended actions. However, she argues that failure does not undermine her claim that our knowledge of intentional action is non-observational.

One of the distinguishing features of one's knowledge of one's own intentional action, according to Anscombe, is that it is practical knowledge. Practical knowledge, as she explains, is "the cause of what it understands," as opposed to theoretical/speculative knowledge, which "is derived from the objects known" (Anscombe 1957: 87). This means that while practical knowledge is influenced by the external world, it also leads to changes in it, and those changes are what comprise the object of this knowledge. The prevailing interpretation of Anscombe's remark is that the fit between such knowledge and its object runs in the opposite direction than it does in cases of theoretical knowledge. The direction of fit in practical knowledge is from the world to the mind, in contrast to theoretical knowledge, which we understand as the mind's adaptation of the external world, with its direction of fit being from the mind to the world.

The difference between practical knowledge and theoretical knowledge can be demonstrated better in Anscombe's well-known example of the shopping list:

Suppose someone goes shopping with a shopping list. The items in his shopping cart and the items on the shopping list should match. If there is a discrepancy between

the items listed and the things purchased, where does the error lie? In other words, if the intention is not consistent with the action, where does the agent's error come in? According to Anscombe, "the mistake is not in the list but in the man's performance. (if his wife were to say: 'Look, it says butter and you have bought margarine,' he would hardly reply: 'What a mistake! we must put that right' and alter the word on the list to 'margarine')." (Anscombe 1957: 56).

Now suppose that while shopping, the person is followed by a detective who records everything he buys in a list. In this scenario, if the purchased items do not match what the detective has recorded in his list, it is no longer a mistake in the detective's performance but a mistake in the written list itself and, therefore, in the detective's judgment about what was bought by the shopper. The shopper's knowledge of his own shopping items is practical knowledge because it is the cause of what it understands. Any discrepancy between the shopping list and the purchased items is due to the shopper's own performance. In contrast, the detective's knowledge about the shopper's shopping is theoretical knowledge because it is derived from the objects known. Any discrepancy between the recorded list and the purchased items is due to an error in the detective's observation or judgment.

In other words, beliefs are mental attitudes that aim at the truth of their content, and if there is a contradiction between the content of a belief and the world, it is the belief that should be revised to fit the world. Intention, on the other hand, is a mental attitude that aims at bringing about a particular state of affairs in the world. If an agent fails to achieve the intended state of affairs, the problem lies in her performance rather than in the intention itself. The world needs to be changed to align with the intention, rather than the intention being changed to align with the world.

To summarise what we have said so far about Anscombe's theory of intentional action, we must say that, according to Anscombe, intentional action is an action to which the question 'why did you do this?' in its special sense has application. Two answers that reject this application are 'I did not know I was doing it' or 'I know through observation that I did it.' Therefore, one must have non-observational knowledge of one's intentional action. Donnellan's view is that failures in doing things lead us to consider one's knowledge of what happens in the outside world through one's action to be observational, but Anscombe rejects this view, stating that the mistake lies in one's performance rather than in one's judgment. Nevertheless, Anscombe still needs to explain how this non-observational knowledge, with a different direction of fit, is epistemically possible. In the next section, we will discuss

Kieran Setiya's defence of non-observational knowledge of intentional action, which is consistent with PTA and has been central to the theory.

3.5. How is such non-observational knowledge possible?

To explain how non-observational knowledge of one's own intentional actions is possible, Setiya initially focuses on the belief that accompanies those actions. According to him,

“Even if it is true that in doing something intentionally one must believe that one is doing it, this belief will not always amount to knowledge.” (Setiya 2008: 389).

However, the very existence of such beliefs is questionable. Setiya reminds Donald Davidson's example, which we mentioned earlier in section 2.4., about a man writing and making ten carbon copies, doing both intentionally, and yet not knowing that he is making copies, but just that he is trying to make them. Then, Setiya adds: “In a case like this, the carbon-copier need not even believe that he is making ten copies, since he doubts that the pressure will go through so many times.” (Setiya 2008: 390). Therefore, the question arises as to whether we can speak of a specific knowledge when we are not even sure if there is a belief accompanying the intentional action.

To clarify the kind of belief necessary for non-observational knowledge of intentional action, Setiya argues that when someone is intentionally engaged in a complex action without being fully convinced of its success, it is still implausible to consider her completely unaware of all her actions, including the less complex ones, taken in order to achieve her goal. For instance, a carbon copier may not be certain if the pressure will go through all the copies, but he still knows that he is pressing on the paper and that by doing so, he is making copies. Therefore, Setiya claims that to do something intentionally is to believe that one is doing something relevant to one's intention. In this regard, Setiya qualifies the claim that intentional actions are accompanied by beliefs:

“If A is doing ϕ intentionally, A believes that he is doing it, or else he is doing ϕ by doing other things, in which he does believe.” (Setiya 2008: 390)

Setiya's qualification implies that in doing something intentionally, there are at least some basic intentional actions, those done without doing anything else intentionally, and that such basic intentional actions are necessarily accompanied by belief. However, a potential challenge to this argument arises in the case of someone who is uncertain about her ability to perform even such a basic action. (Paul 2009: 553). What if the carbon copier is even uncertain of his power to move his hand and yet manages to do it? Imagine a case of periodic partial paralysis. She may occasionally

fail to move as she desires but she cannot claim that all her movements are henceforth unintentional. It would also be difficult for her to pretend that she is moving her hand by doing other things in which she believes. So, again, how can she know that she is moving her hand if each time she manages to do it she cries out: 'Great! I could not believe it!'

For Setiya, this question is a result of simplifications made in epistemology and will not be raised if we stop "ignoring the fact that belief comes by degree." (Setiya 2008: 391). In the case of the carbon copier with periodic partial paralysis, one is likely not thinking of one's movements in terms of unequivocal belief or disbelief, but rather a level of confidence. Setiya's point is that if someone can perform an intentional action without believing that she is doing it (or even that she is taking the means to achieve that end), then it would also be possible for her *to perform an intentional action and believe that she is not doing it*, or even that she is not taking any actions that lead to that end. This sounds really odd, and to prevent such consequences, Setiya proposes to consider the possibility for intentional actions, including basic ones, to be accompanied by partial beliefs. Hence, the above qualification is reformulated to incorporate the idea that belief comes by degree.

"If A is doing ϕ intentionally, A believes that he is doing it or is more confident of this than he would otherwise be, or else he is doing ϕ by doing other things for which that condition holds." (Setiya 2008: 391)

Hence, according to Setiya, there can be no intention without at least some accompanying belief, even if it is in a basic action. However, this raises the question of how one can be justified in forming beliefs that intentions involve. This is a difficult task, as if forming an intention is indeed forming a belief, what distinguishes this belief from mere wishful thinking? For example, could the carbon copier be justified in forming a belief that he intends to simultaneously make one hundred copies? The necessary presence of belief in intentional action rules out the possibility for it to be formed by some contingent mechanism. This leads to a puzzling situation because the transition from intending to forming a belief about what one is doing or going to do should be warranted, otherwise it would be wishful thinking. To show how Setiya solves the problem, let us go back to where he begins his argument:

"Even if it is true that in doing something intentionally one must believe that one is doing it, this belief will not always amount to knowledge. Suppose that I am clenching my fist intentionally, and in doing so I believe that I am clenching my fist. In order to know that I am clenching my fist, I must know that I am able to clench my fist, in the

simple conditional sense: I must know that, if I intend to be clenching it, I will be doing so in fact." (Setiya 2008: 389)

This knowledge of ability, as Setiya calls it, is a necessary condition for the transition to be warranted. Let's revisit the case of periodic partial paralysis. Suppose that the individual has become tired of her condition and, in a fit of anger, irrationally believes that she is cured. She intends to clench her fist and manages to do so, not because she is cured, but because her paralysis is not permanent. In this case, the question is not about her confidence level while performing the intended action, but rather whether she has good reasons to believe she is capable of doing it. According to Setiya, even if she now believes to a higher degree that she is clenching her fist, she is still not justified in forming that belief.

According to Setiya, simply having the knowledge of ability is not enough to justify beliefs in intentional actions (Setiya 2008: 393). While it is true that someone in good health and with freedom of movement can confidently expect to clench her fist if she intends to do so, this alone does not provide a reason to believe that she is indeed acting on her ability. Let us not forget that, for Setiya, one's intention to clench one's fist involves a belief about what is going on when one clenches her fist. What is going on is that she is acting in a certain way and she believes that she is acting that way because she knows how to clench her fist. According to Setiya,

"Knowing how to ϕ is the state or condition that, with knowledge of ability, provides the epistemic warrant for decision. Together, they justify the transition in which one forms the intention and belief that one is doing ϕ or that one is going to do it." (Setiya 2008: 407)

Setiya is saying that knowledge of ability alone is not enough to justify belief in intentional actions, but when combined with knowledge-how, it provides the necessary epistemic warrant for believing that one is doing or going to do the intended action. To explain more, Setiya invites us to consider prospective intention (Setiya 2008: 406). Suppose a couple decide to dance the tango at their wedding. This decision seems unjustified unless they know already how to dance the tango. Otherwise, they can only claim that they have decided to learn how to dance the tango and, if the learning process is a success, exercise the acquired knowledge at their wedding. In other words, they are entitled to believe that they are going to dance the tango only when they have the needed knowledge-how. Moreover, as soon as they have obtained the knowledge of how to dance the tango, they are justified to believe that they are going to act on this ability at their wedding. The same reasoning, Setiya argues, can

be applied to intentions in acting. (Setiya 2008: 406). The knowledge of how to dance the tango is a persisting state, of course if one exercises it from time to time. Married for better or worse, our exemplary couple can now dance the tango whenever they intend to, each time forming the warranted belief that they are doing so. The reason is not just that they are now aware that they are able to, but that knowing how to do so justifies them in believing that they are acting the way they should.

Here, Setiya presents a case of complex actions that can be problematic (Setiya 2008: 404). Setiya gives the example of a person trying to defuse a bomb. Even if the person is an expert, she could be faced with an original and sophisticated apparatus that she does not know how to deal with. The problem becomes more significant if she must defuse the bomb just because she has arrived at the wrong place at the wrong time. Despite the complexity of the situation, the person intends to defuse the bomb and decides to disconnect a red wire. As a result, the bomb stops. The question is whether she intentionally defused the bomb. Setiya's answer is yes because, even though the person may have been lucky to disconnect the right wire, she knew how to take the relevant means. In other words, she meant to defuse the bomb by deciding to disconnect a wire. Therefore, a similar qualification to the one proposed by Setiya for belief in intentional action can be formulated for knowledge-how.

“If A is doing ϕ intentionally, A knows how to ϕ , or else he is doing it by doing other things that he knows how to do.” (Setiya 2008: 404)

Similar to the previous one, this qualification identifies at least basic intentional actions (those done without doing anything else intentionally) as actions for which one must exercise one's related knowledge-how and thus accompanied by justified beliefs. Considering that our competencies could be as partial as beliefs we have in them, the whole argument may be paraphrased as there can be no intentional action without the exercise of at least knowledge-how of some maybe basic actions that are relevant means for doing the intended action.

As we mentioned earlier, Setiya's idea regarding knowledge of intentional action has had a substantial impact on the development of PTA. In the following section, we will compare Setiya's idea of intentional action with PTA to explore their similarities and differences.

3.6. Setiya's account and PTA

As we saw in the last section, in his article 'Practical Knowledge', Setiya distinguishes between knowledge-how and knowledge of ability, and argues that both are necessary and sufficient for justifying the beliefs involved in doing something intentionally. However, in his later article 'Practical Knowledge Revisited,' he reconsiders the necessity of knowledge of ability, suggesting that "perhaps it is sufficient that one knows how to ϕ and has no reason to doubt one's own ability." (Setiya 2009: 134).

The challenge arises from the possibility that one may know how to do something, yet be unable to actually do it. For instance, one may know how to move her arm, but if it is tied to a chair, she will be unable to do so. In the previous chapter, we presented an example of a swimming champion who had recently lost an arm. It was noted that there is a difference between knowing how to swim with two arms and knowing how to swim with one arm. The champion may know the former, while she does not know the latter. Furthermore, we said that knowledge-how comes in degrees. Thus, we can say that the champion knows much better how to swim with two arms than with one arm.

Now, consider the situation in which our swimming champion will not be able to swim at all. Suppose she travels to a planet where there is not enough water for her to swim. In this case, can she intend to swim? Those who answer this question in the affirmative should reflect on the difference between intention and desire. Intending to do something requires a commitment to carry out that action, while desire does not necessarily entail commitment. One can have the desire to ϕ in spite of whether one knows how to ϕ , since having a desire to do something does not bring any commitment to do that action. However, intending to ϕ requires a commitment to ϕ and doing something in order to do ϕ .

If a person cannot swim at all, it can be said that there is no degree of knowledge of how to swim that she can exercise in order to swim. Therefore, in the case of our champion who has travelled to another planet, we cannot say that she intends to swim, while she may have the desire to swim. As a result, it seems that Setiya's doubt about the need for knowledge of ability is reasonable, and for a person to have a justified belief that 'she is ϕ -ing/going to ϕ ', it is sufficient that she knows how to ϕ .

Furthermore, Setiya's solution to the problem of lack of knowledge in how to do some complex actions is to require that the agent "must at least know how to take some relevant means." (Setiya 2008: 404). These relevant means involve at least some basic actions, which the agent certainly knows how to perform. However, merely knowing how to perform basic actions or even simple actions like cutting a wire is not a sufficient justification for believing that one is successfully completing a complex operation like defusing a bomb. This is because it is unclear how these "relevant means" are related to the intentional action that, as Setiya says, is "not accompanied by knowledge how." (Setiya 2008: 404).

As we mentioned in the previous chapter, when one is ϕ -ing intentionally, according to PTA, one must know to some extent how to ϕ and exercises this knowledge. As we said, the qualification 'to some extent' could be used in cases where the agent does not know how to ϕ in the first place but tries to do it by doing something else she knows how to do. Nevertheless, choosing other things that one knows how to do, in order to ϕ , is not arbitrary. Instead, these are things that the agent believes that if she were to ϕ , it would be in that way.

Regarding intentionally defusing a bomb, Setiya says that "I know how to cut the red wire, and I think it might defuse the bomb, even though I can't be sure." (Setiya 2008: 404). How can one think that cutting wire might defuse a bomb, without knowing how to defuse it? One possible explanation is that the person had watched some police detective films before, or heard in the news that a bomb disposal technician managed to prevent the bomb from exploding by cutting the wires, or maybe this person had noticed that the bomb was connected to a battery through a wire and has argued that if we disconnect any electrical appliance from its battery, it will stop. In any case, the fact that the person has this belief indicates that she has to some extent knowledge of how to defuse a bomb, even if it is not complete or expert-level knowledge.

Thus, it follows that knowledge-how can come in degrees, as a bomb disposal technician would know more about defusing bombs than an amateur agent. As a result, it can be said that in the case of an intentional action that one does not quite know how to do, Setiya rightly pointed out that the agent must at least know how to take some relevant means. However, this does not mean that the agent does not know anything about how to do it. Rather, the fact that the agent must at least know how to take some relevant means is a good reason to say that the agent knows to some extent how to do it.

4. Three Versions of the Transparency Account of Intention

[T]here is a sense in which he unfailingly knows what he is trying to do, in contrast with an observer, simply because it is his intention and not anyone else's. There is no question or possibility of his not knowing, since doing something with intention, or intentionally, entails knowing what one is doing; and intending to do something on some future occasion entails already knowing what one will do, or at least try to do, on that occasion. There is therefore no need of the double, or reflexive, knowing which would be implied by the cumbersome phrase 'knowing what one intends'. To say 'I know now what I intend to do' is a redundant way of saying 'I know now what I shall do', and 'I know what my intention is in doing this' is an impossibly redundant way of saying 'I am doing this with intention or intentionally'.

(Hampshire 1982: 102)

Abstract

In this chapter, we assess Moran's, Boyle's, and Byrne's transparency accounts of intention and illustrate the challenges they encounter in furnishing an epistemology of intention. Furthermore, we emphasise how PTA can expose their inadequacies and, on some occasions, assist them in surmounting these obstacles.

4.1. Introduction

Evans believes that one's own belief is transparent to the fact that the object of this belief is the case. Although an adequate explanation of transparency is necessary to elucidate how this transition from the world to the mind occurs, it appears more plausible that our beliefs are transparent to the world than our intentions, which have always seemed resistant to this perspective. The divergent rules governing the direction of fit for intention and belief shed light on this disparity.

Anscombe's renowned example of the shopping list highlights that beliefs aim at truth, and if we discover that something in the world contradicts our belief, that means the problem is in our judgment, and we should modify our belief to conform to the world, not the other way around. Conversely, an intention is a mental state that necessitates realisation in the world. If the individual fails to execute her intention in the world, the intention should not be altered to fit the world's circumstances; instead, the issue lies with the individual's performance, and the world should be altered accordingly. As we mentioned earlier in section 3.4., generally, intentions and beliefs have distinct "directions of fit" between the mind and the world - the direction of fit in intention is world-to-mind, whereas in belief, it is mind-to-world. Consequently, by turning our attention outward in the case of beliefs, we abide by the rules concerning the direction of fit, and the world-to-mind pattern of the transparency account appears to be appropriate. However, following these rules, if we shift our focus outward in the case of intention, we would be unsuccessful in comprehending what our intention is.

Additionally, it seems that the different directions of fit correspond strongly with different types of reasoning and, thus, with different forms of knowledge - theoretical and practical. In theoretical reasoning, one is looking outward to ascertain what is true, and the outcome of this inquiry is a belief about external phenomena. However, in practical reasoning, one is inquiring about what one should do, and the result is an intention or an action that generates a change in the external world. The nature of belief is such that it aims to conform to a rational thinker's assessment of what is true or what is best supported by evidence. However, it is not evident that intentions can be transparent in the same way. Even in an ideally rational agent, it is unclear whether there is a single question or fact to which intentions are transparent. (Paul 2014: 301)

Although the distinct nature of intention appears to weaken the transparency

account's ability to clarify how individuals access their own intention, some authors have endeavoured to uphold the effectiveness of the transparency account for intention in recent years. In this chapter, we will evaluate Moran, Boyle, and Byrne's distinct versions of the transparency account of intention individually and demonstrate that their accounts do not furnish a proper epistemology for intention. Additionally, the potential of PTA to reveal their shortcomings and aid in overcoming some of the obstacles will be highlighted.

4.2. Moran

4.2.1. Introduction

In the first chapter, we presented Moran's Transparency account. We said that, according to Moran, individuals have privileged access to their mental states, meaning that access to their minds is both immediate and authoritative. There, we attempted to outline Moran's response to the problem of self-knowledge, which poses the question of how one can make authoritative assessments of one's own mental life when these assessments are not founded on evidence, observation, or inference.

As we said in the first chapter, in response to this question, Moran differentiates between the deliberative and theoretical stances that individuals can adopt toward themselves when addressing questions about their mental states, such as "What do you believe?" or "What do you want?". According to Moran, individuals can typically make authoritative assessments of their attitudes without relying on any evidence, observation, or inference because they generally adopt a deliberative stance towards themselves. We clarified that adopting a deliberative stance towards oneself involves actively participating in shaping one's own attitudes. Regarding this active role, Moran asserts:

The primary thought gaining expression in the idea of 'first-person authority' may not be that the person himself must always 'know best' what he thinks about something, but rather that it is his business what he thinks about something, that it is up to him. In declaring his belief, he does not express himself as an expert witness to a realm of psychological fact, so much as he expresses his rational authority over that realm. (Moran 2001: 123-4).

Moran posits that since our thoughts, beliefs, and intentions are our own business, it is our responsibility to adopt a rational attitude in response to questions like "What do I believe?" According to Moran, our mental attitudes are transparent to the world because, when forming our rational attitudes, we direct our attention "at least equally towards the outward," towards the object of our response, as towards ourselves. (Moran 2001: 59). Thus, in response to questions such as "What do I believe?", we are not turning our focus inward passively to identify our belief. Instead, our answer is derived from actively directing our attention to the world and forming our belief

based on the evidence available to us.

Regarding the transparency of beliefs, Moran contends that “with respect to belief, the claim of transparency is that from within the first-person perspective, I treat the question of my belief about P as equivalent to the question of the truth of P.” (Moran 2001: 62-63). As we highlighted in the first chapter, Moran maintains that having sufficient reasons for P obliges us not only to the truth of P but also to the truth of our belief about P. Brueckner summarises Moran's stance by stating that for Moran, if one is a rational agent with a genuine system of beliefs, then one must adopt a deliberative stance towards one's own beliefs, whereby the question ‘Do I believe that P?’ is transparent to the question ‘Is P true?’. (Brueckner 2003: 196).

An example of the transparency of beliefs in the way Moran presents it can be illustrated through a scenario where an individual observes raindrops on her window and hears the sound of rain falling outside. Based on this evidence, she concludes that it is indeed raining outside. If someone were to ask her, "Do you believe that it is raining outside?", she could confidently respond with "Yes, I believe so." In this case, the individual is able to attribute the belief to herself with authority, without the need for additional evidence, as she has sufficient evidence to support the belief that it is raining outside.

In this section, our objective is to exemplify how Moran employs his approach to the self-knowledge of intention and to identify its limitations. In the subsequent subsection, we will endeavour to employ Moran's concept of transparency to the mental state of intention.

4.2.2. Moran's Transparency account of Intention

Moran makes a distinction between how one engages with one's own intention and how others confront it from a third-person point of view. According to him,

“... in declaring the intention the person is committed both to the practical endorsement of the action and the expectation of a future event. And the hearer is thereby told something about the future which he may doubt, or count on like a prediction.” (Moran 1997: 153)

We can say that this distinction aligns with Moran's differentiation between theoretical and deliberative stances. From a third-person point of view, one takes a

theoretical stance on someone else's intention and thus is not involved in the practical endorsement of the action or the expectation of a future event. Here, Moran emphasises the unique manner in which one engages with one's own intention from a deliberative/first-person stance.

Elsewhere, with respect to knowledge of one's own intentions, Moran draws attention to a distinction between two kinds of certainty with respect to one's own intentions. One kind of certainty is based on evidence or discovery, while the other kind is based on a decision made by the person. (Moran 2001: 55). According to Moran, any uncertainty about what one intends to do is typically a result of not having fully formed the intention. (Moran 2001: 55-6). In other words, uncertainty about one's own intention cannot be eliminated by simply observing one's own intention—in a way that one finds out somebody else's intention—and saying that 'Aha! That is what I intended to do!'. Instead, one's uncertainty is eliminated by making a decision about what to do. (Moran 2001: 56). Regarding these two distinct perspectives that one can adopt on one's own intention, Moran says:

My knowing what I will do next is not based on evidence or other reasons to believe something, so much as it is based on what I see as reasons to do something. Hence, a person's statement of intention is not to be challenged by asking for his evidence. When I make up my mind about what to do, and tell someone else, I do indeed provide him with a reason to expect something, a very good reason if I'm not too vacillating, or a liar; but what I possess myself is not an expectation, based on evidence, but an intention, based on a decision. (Moran 2001: 56)

"Knowing what I will do next" can also refer to knowledge of one's own intentional action. However, in this quote, Moran is specifically referring to knowledge of intention, which he claims is not based on evidence. According to Moran, one can declare one's intention when one make up one's mind and make a decision. Nonetheless, throughout his work, Moran maintains an intimate connection between one's knowledge of intention and knowledge of intentional action. He draws on this connection to explain the "first-person immediacy and authority" (Moran 2001: 124) that individuals enjoy with regard to their own mental states, as well as to clarify how individuals form an intention and become aware of it. (Moran 2001: 56&127).

With regard to knowledge of one's own intentional action, Moran adopts Anscombe's view. In the second and third chapters, we extensively discussed Anscombe's account of intentional action and knowledge of intentional action. We said that according to her, an intentional action is an action "to which a certain sense

of the question 'why?' is given application; the sense is of course that in which the answer, if positive, gives a reason for acting." (Anscombe 1957: 9). Thus, one has intentionally performed an action if and only if one is able to give a particular kind of answer to the question "why did you do that?" or "why were you doing that?" that provides reasons without which the action would not have been intentional. We also noted that, according to Anscombe, one knows what one is doing intentionally, and that this knowledge is non-observational.

Moran formulates Anscombe's idea as: if a person knows her own reasons for adopting the aim to ϕ , she knows that she is ϕ -ing. About the connection between one's knowledge of one's own reasons for doing an action and one's knowledge of one's own intentional action, Moran says:

"For an agent to conceive of himself as capable of forming an intention and implementing it (which, I take it, is necessary to conceiving of himself as an agent at all) he must take his intentional action to be determined by his reasons, and thus he is in a position to know a true description of his action in knowing his reasons." (Moran 2001: 127).

Therefore, it can be said that, for Moran, one knows what one is doing based on one's knowledge of one's own reasons for doing that action, and that these two types of knowledge together lead to the formation of one's intention. It is important to keep in mind that, for Moran, individuals typically adopt a deliberative stance toward their own mental states, and play an agential role in forming their own attitudes. When Moran refers to the agent as someone "capable of forming an intention and implementing it," his emphasis is on adopting a deliberative standpoint towards one's intention. As Owens describes Moran's view, "one is an authority about what one is doing and why just insofar as one is occupying this standpoint. To occupy this standpoint is, by definition, to both make judgements about what you have reason to do and to implement those judgements in decision and action." (Owens 2003: 797). In other words, one has authority over one's own actions and intentions to the extent that one can deliberate on one's reasons for action and then act accordingly.

In our previous discussion in the first chapter, we explored Moran's Transparency account of belief and his claim of transparency: "the claim of transparency is that from within the first-person perspective, I treat the question of my belief about P as equivalent to the question of the truth of P." (Moran 2001: 62-63). Now if we want to provide a picture of Moran's Transparency account of intention and explain how, for him, directing our attention "towards the outward" can explain the immediate and

authoritative access we have to our intentions, we can say, with respect to intention, 'the claim of transparency is that from within the first-person perspective, I treat the question of my intention to ϕ as equivalent to the question of my reasons for doing ϕ . Thus, by having sufficient reasons for doing ϕ , I am committed not only to the practical endorsement of the action ϕ , but also to the truth of my intention to ϕ .'

4.2.3. Objections

As discussed in the previous section, Moran's Transparency account of intention connects an individual's knowledge of her own intention to two elements: 1) the individual's knowledge of her own intentional action and, 2. having reasons for doing the action or knowledge of the reasons for performing the action. However, linking knowledge of intention with each of these elements gives rise to certain issues.

We begin by presenting the problem of connecting knowledge of intention with knowledge of intentional action, in the way that Moran depicts this connection. It is not clear what the difference is between knowledge of one's own intention and knowledge of one's own intentional action in Moran's discussion. As mentioned in the previous subsection, Moran sometimes talks about "knowing what I will do next," which can refer to both types of knowledge. Even if we assume that Moran sees these two types of knowledge as distinct, his emphasis on the deliberative position and the active role of the self-knower suggests that knowledge of one's own intentional action leads to the formation of one's own intention, rather than to knowledge of it. It is not clear in Moran's work how Transparency explains the transition from knowledge of one's own intentional action to knowledge of one's own intention. Two philosophers who refer to this issue in Moran's account are Lucy O'Brien (2003) and Sidney Shoemaker (2003). According to O'Brien:

The difficulty in determining how Moran thinks rational agency gives us knowledge of its product comes to a significant extent from there being little guidance as to what positive conditions he places on knowledge. We know that Moran does not take traditional forms of justification by observation as required for knowledge, and it is clear that he thinks mere reliability is insufficient for knowledge, at least in this case. However, there is less guidance as to what is required. (O'Brien 2003: 380)

If we accept Moran's thesis, when one intends to do something, one knows one's

intention by exercising one's rational agency and knowing one's intentional action, while someone else knows it by observing one's actions. Both parties know the same thing through different means. However, while the other's belief about the intention is justified by observation, what justifies the self-knower's belief in what she herself intends to do is unclear. How can the self-knower, through knowing what she is doing, self-ascribe her intention in an epistemically secure way when this self-ascription is not based on observation? Sidney Shoemaker also raises a related difficulty, arguing that Moran's account "in terms of epistemic agency may explain how we get to the affirmation or reaffirmation of the content proposition, but it seems to me not to explain how we get to the affirmation of the self-attribution." (Shoemaker 2003: 397). Even if we assume that Moran is correct and that one must answer the question of 'what to do?' by reflecting on one's reasons for action, it only leads to an affirmation of what must be done. With respect to belief, Shoemaker explains:

"If reflecting on the reasons is just reflecting on facts (or presumed facts) about the world that are (or would be) evidence that P, this can certainly contribute to his arriving at an affirmation that P, but it has yet to be explained how can it lead to his arriving at an affirmation of 'I believe that P'." (Shoemaker 2003: 398)

Moran attempts to address the objections raised by O'Brien and Shoemaker in his 2003 article. He frames their challenge as the following question: "What right have I to think that my reflection on the reasons in favor of P (which is one subject-matter) has anything to do with the question of what my actual belief about P is (which is quite a different subject matter)?" (Moran 2003: 405). Moran's response to this challenging question is:

When I say that what I believe is up to me, I mean that, unlike the case of sensations or other non-intentional states, I take what I believe to be answerable to my sense of reasons and justification, and I take myself to be responsible for making my belief conform to my sense of the reasons in favor or against. And now to bring this back to the Transparency claim, my thought is that it is only because I assume that what I actually believe about X can be determined, made true by, my reflection on X itself, that I have the right to answer a question about my belief in a way that respects the Transparency Condition. (Moran 2003: 406)

In this regard, through the Transparency method, one can self-ascribe one's intention in an immediate and authoritative manner insofar as one has the right to assume that one's intention is determined by one's reflection on one's reasons. (Gertler 2011a: 188). However, this raises the question of whether it is permissible to assume that one's intention is determined by such reflection. Here we come to another type of

objection that is raised more due to the connection between knowledge of intention and having reasons to do the action. This objection concerns a potential gap between reason and action, as seen in akratic actions. Some philosophers argue that one's intention cannot be determined by reflection on one's reasons, as one may have a good reason to do something but refrain from doing so, or intend to do something without having any reason to do it. As Jonathan Way explains:

[A]gents don't always have the attitudes that they judge are the attitudes to have. I can intend to have another drink, while knowing full well that this is not the thing to be doing...can I not know – in the ordinary, first-person way – that I intend to have another drink, even though this attitude is not transparent?(Way 2007: 226).

According to Way, akratic attitudes are rational attitudes which can uncontroversially be the objects of ordinary self-knowledge, but Moran's account of Transparency is incapable of explaining the privileged access that one has to this type of attitude.

This study disagrees with the idea that, in determining one's intention, thinking about one's reasons plays no role. We think that having reasons for doing something plays a crucial role in forming an intention and carrying out the action, as well as in knowing one's intention and action. However, we concur with critics who argue that thinking about reasons alone does not grant authority to self-ascribe an intention. We need something more, and that is the agential role one plays in committing to carry out an action. Moran's ideas align with this perspective, as he emphasises the deliberative role that individuals play in knowing their own attitudes. However, since he did not clearly explain the transition from world to mind in his account of Transparency, this paves the way for such objections. The following discussion will attempt to explain why having reasons for doing something is important in the formation and knowledge of intentions.

Having strong reasons for or against doing an action but intending to do the opposite is similar to cases in which one knows one's own reasons for adopting the aim to ϕ , but does not ϕ intentionally. This problem was previously discussed in Chapter 2 in the context of the rational theory of action and its difficulties in cases of causal waywardness, where one has a reason for doing something, intends to do it, and will do it, but not intentionally. The case that we presented as an example there was as follows: John is a virologist working on COVID-19 who intends to transmit the virus to Sarah, who is suffering from fatal cancer and has requested him to infect her in the hope that it could help John find a cure for COVID-19. Unbeknownst to John,

he has already passed on COVID-19 to Sarah on a previous visit. Sarah dies of the virus just one second after John transmits it again.

Accordingly, the objection to the rational theory of action is that John had a reason to kill Sarah through the transmission of the Coronavirus, and he intended to do so, yet he did it unintentionally. Therefore, it is not right to say S intentionally V-ed at T if and only if S V-ed at T and S V-ed for a reason. However, we argued against this objection in the second chapter, as we did not find it valid. The reason for rejecting this objection was that although John transmitted the disease to Sarah, it was not transmitted for the reason he had for the transmission. We explained that although John knew how to transmit the virus and caused Sarah to die, he did not supervise the operation in light of the reasons he had for carrying out this action. Based on PTA, Sarah's death happened unintentionally because the exercise of John's knowledge-how did not cause it. The transmission of the virus was not caused by the exercise of John's knowledge-how since the transmission was not supervised in light of the reasons he had for carrying out the action.

In this case, if the operation was not supervised at all by John, it can be concluded that Sarah's death was unintentional. The failure in supervising the operation also raises doubts about whether we can say that John intended to transmit the disease to Sarah. We can concur with Moran that if someone intends to do something, that intention brings commitment. As we explained in detail in the second chapter, this commitment manifests itself in the supervision of the operation. If John was not yet committed to doing the action, it is difficult to say that he intended to transmit the virus. When we want to talk about John's attitude, we have to stay with *desire*.

The similarity between the Way and John cases lies in the role of reasoning in forming a relevant intention and intentionally doing a pertinent action. In akratic actions, if the action is done intentionally, the agent is exercising her knowledge-how to carry out that action. In doing this intentional action, the agent supervises the pertinent operation, and this supervision must be in line with the reasons she has for carrying out the action. Otherwise, we may question why the agent is supervising this particular operation but not another. Although, in akratic cases, one has a strong reason for doing/not doing an action, one's intentional omission/action is supervised in light of other (maybe weaker) reasons one has for one's omission/action. Therefore, instead of talking about having a reason in general, we should focus on having a reason to do a particular action. Way may have many good reasons for not having another drink, but the fact that he intends to have another drink does not mean he has

no reason for his action. He may be thirsty or want to accompany his friends, and as long as he drinks intentionally and supervises his action, he has reasons for doing so, and in light of those reasons, he supervises his action.

Another problem Way raises concerns the reasonings that lead to two different options, neither of which has precedence over the other. In some cases, one's reasoning about whether to do M or N does not result in a clear answer, because the individual does not see a priority in choosing one over the other— both are equally attractive and equally effective as a means to achieve the goal. According to way,

This suggests that transparency will only work in general on the assumption of uniqueness: given a set of reasons R, bearing on S's having some attitude M, either S ought to have M, or S ought not to have M. But it is widely agreed that uniqueness does not hold for intentions. (Way 2007: 226).

Similarly, we cannot say that, since reasoning does not lead to a more preferable option, one's reasoning plays no role in forming one's intention. In such cases, we must specify the intentional action we are considering. What is the operation we are trying to supervise? When someone goes to a supermarket and chooses a tomato can out of twenty similar cans, if you ask her 'why that can?', she may or may not give you a reason. The question could be asked with regard to the intention to buy a can of tomato in general or to buy this particular tomato can. Suppose the shopper did not find herself committed to purchasing that particular can of tomato. In that case, it does not seem her action of buying that specific can of tomatoes was intentional and she intended to buy that can. Nevertheless, we can still say that the action of buying a tomato can was intentional, and the agent had an intention to buy a can, as long as she was committed to doing so. We should keep in mind that, as we mentioned in chapter two, there is no need to divide actions merely into intentional and unintentional, rather we can assume that there is another category of action as non-intentional.

4.3. Boyle

4.3.1. Introduction

The primary objective of this section is to demonstrate how Boyle's transparency account can elucidate the self-knowledge of intention, and to highlight the limitations of Boyle's perspective when extended to intentions. As we noted in the first chapter, Boyle's Transparency account is founded on the reflective approach, which is rooted in Moran's account. In contrast to what he calls the epistemic approach, the reflective approach is not concerned with identifying a distinct method of arriving at a mental attitude. Rather, it seeks to explain how normal knowledge of one's own mental state reflects something about what it is to have a mental attitude. (Boyle 2011: 235).

As previously mentioned, according to Boyle, when one has a mental attitude and can express it, for one's expression to serve as an indicator of one's knowledge of one's own mental state, one must understand the sentence used to express one's mental attitude. To comprehend a sentence, one must understand its various components. When expressing one's mental state, understanding the different elements of the expression necessitates recognising the mental state mentioned as a certain condition one is in. Therefore, when one expresses one's mental state, one apprehends what being in that certain mental state means. In subsection 1.3.4., we formulated this condition on self-knowledge as below:

- **The Minimal Condition on Self-Knowledge (MCS):** When someone expresses the mental state that she is in, she must understand whatever sentences she uses for this expression, in the sense that her expression must be a *representation* of her own *certain* condition.

Subsequently, we discussed how Boyle addresses the question of what allows individuals to understand the content of their own utterances, given the MCS. Boyle posits that, in general, to comprehend the content of an utterance regarding their own mental state, individuals must be able to reflect on relationships between the content of any given sentence about their own mind and the content of various other sentences, in a way that they can recognise which sentences are true. According to Boyle, to comprehend and acknowledge a claim as true, one must reflect on one's

grounds for making that claim. Boyle asserts that "a comprehending speaker must be able to entertain a certain sort of "Why?"-question about the claims she makes, a question that asks for grounds that show the claim in question to be true." (Boyle 2009: 151). Taking these points into account, we formulated Boyle's idea as follows in section 1.3.5.:

- **The Ability to Understand (AU):** To meet MCS, one must be able to reflect on relationships between the content of any given sentence about one's own mind and the content of various other sentences, in a way that one can recognise, based on one's grounds, which sentences are true.

We said that according to Boyle, an individual who possesses the ability to self-represent oneself in the manner that he described is entitled to accompany one's expression of one's own mental state with the phrase 'I believe.' In Boyle's view, the question 'Do I believe that p?' is transparent to the question 'Is p true?', because if someone can sincerely claim that P is the case, after deliberating whether p, one's understanding of one's claim entitles them to say "I believe that p."

Even if we assume Boyle's account is valid with respect to our knowledge of beliefs, desires, and emotions, it is uncertain if his approach is adequate for intentions. In the following subsection, we will explore whether Boyle's account of transparency can be extended to the mental state of intention.

4.3.2. Boyle's Transparency account of Intention

According to Boyle's Transparency account, understanding what intending to ϕ means is a requirement for self-ascribing the intention to ϕ . Therefore, we can revise MCK as follows:

- **The Minimal Condition on Knowing One's own Intention (MCKI):** When one expresses one's own intention to ϕ , one must understand what having the intention to ϕ means and can identify oneself as the person who intends to ϕ .

Moreover, Boyle's account suggests that meeting MCKI requires satisfying the following condition:

- **The Ability to Understand One's own Intention (AUI):** one must be able to reflect on relationships between the content of "I intend to ϕ " and the content of

various other sentences, in a way that one can recognise, based on one's grounds, which sentences are true.

To gain a better understanding of MCKI and its requirements, it would be helpful to consider Anscombe's famous remark on the expression of intention that was mentioned earlier in section 2.3.. There, we outlined that Anscombe distinguishes three different concepts that can be employed by the term 'intention.' Firstly, intentions could be revealed by expression of intention/intention for future ('I will do something'). Secondly, they could also be worked out by an intentional action/intention in action ('I am doing something'). Lastly, intentions could be disclosed by an intention with which actions are done ('I am doing something in order to do something else'). Accordingly, one can reveal one's intention to ϕ by expressing, "I will ϕ ," "I'm ϕ -ing," or "I'm ϕ -ing in order to do something else." For the sake of simplicity, we assume the second one and reformulate MCKI and AUI:

- **The Minimal Condition on Knowing One's own Intention* (MCKI*):** When one expresses "I'm ϕ -ing," one must understand what doing ϕ means and can identify oneself as the person who is doing ϕ .

- **The Ability to Understand One's own Intention* (AUI*):** one must be able to reflect on relationships between the content of "I'm ϕ -ing" and the content of various other sentences, in a way that one can recognise, based on one's grounds, which sentences are true.

So far we can say, according to Boyle's account, for someone to intentionally do ϕ , she must understand what it means to do ϕ and be capable of reflecting on the relationships between the content of any given sentence regarding ϕ -ing and the content of various other sentences. Additionally, one should be able to distinguish between true and false sentences concerning ϕ -ing, based on one's grounds for truth or falsity.

The person who has the ability to represent her own intention, in the way that is described, according to Boyle's view, is entitled to use the phrase 'I believe' when expressing her intention. Thus, the question 'Do I believe that I'm ϕ -ing/will ϕ ?' is transparent to the question 'am I ϕ -ing/will I ϕ ?'. This is because when deliberating on what she is doing or what she will do, if she can sincerely claim that she is or will be doing something, then her understanding of her own intention entitles her to say, "I believe I am ϕ -ing/I will ϕ ".

4.3.3. Objections

One objection to Boyle's arguments is that it is not clear how understanding the expression of one's own action leads to first-person self-knowledge about one's intention. In his account, Boyle proceeds in the reverse direction of Moran's and starts from the avowal. It seems that what authorises him to start from the expression of self-knowledge is that he wants to explain Moran's account. However, Boyle not only wants to defend Moran, but also wants to say that the true approach to the problem of self-knowledge is his reflective approach, which we can now summarise as follows: If one understands, albeit in a particular way described by Boyle, the expression of one's own mind, one is entitled to believe that one is in a particular mental condition. However, it is not clear how: 1. this entitlement counts as knowledge, and 2. this entitlement accounts for the immediacy and authority that exist in self-knowledge.

Assuming that the self-knower understands the expression of what she is doing on the basis of the reasons she has for her action, it is not clear how the self-knower can say that she knows what her intention is. In other words, it is not clear how Boyle explains the transition from understanding one's action to knowing one's intention. Boyle may face the same objections that Shoemaker and O'Brien have raised against Moran's account mentioned in 4.2.3.. In his paper *Transparency and Reflection*, Boyle explains the transition as below:

"Now consider a person who makes the transition from

(1a) I will ϕ

to

(2) I intend to ϕ .

The reasonableness of this transition is evident. A person who soundly thinks (1a) already thinks of her future ϕ -ing in a way that implies a present intention to ϕ : her judging (2) just makes this implication explicit. What she must understand in order justifiably to make the transition from (1a) to (2) is simply that the way of thinking of her future involved in (1a) implies a present intention to ϕ . But this is to say that she does not need any further information about her present psychological state beyond what is already contained in (1a). All she needs is a grasp of this condition of application of the concept intention itself." (Boyle 2019: 21)

Although here Boyle is trying to formulate his idea in different terms, the main idea is unchanged: understanding one's expression of intention leads to self-ascription

of the intention. If one understands one's expression 'I will ϕ ', one is justified in believing that one has the intention to ϕ . What makes this belief justify is that, according to Boyle, 'I intend to ϕ ' is implicit in 'I will ϕ '.

It is evident that one's intention is implicit in one's expression of one's intention. However, the challenge is to explain the privileged access one has to this implicit information. Even if there is a reliable epistemic way for the transition from understanding one's own action to knowing one's own intention, it is unclear how this transition can occur in an immediate and authoritative manner.

Moran's account clearly addresses the immediacy and authority that exists in self-knowledge, by emphasising the agential role one plays in forming one's own mental state, which cannot be played by another. However, as we mentioned above, Boyle not only defends Moran's viewpoint, but also introduces the reflective approach to account for self-knowledge. In his account, it may seem that self-knowledge is no different from another's knowledge of our mental states, if the other person also understands our expression of mental attitudes in such a way that she can recognise, based on her own grounds, whether our expression of our mental attitude is true.

For example, one may ask: does understanding one's statement about one's own action demand knowledge of one's own intentional action? Can one understand, in the way Boyle suggests, one's own expression 'I am ϕ -ing/will ϕ ' without knowing that one is ϕ -ing/will ϕ ? It seems odd that one can recognise, on the basis of one's grounds, the truth of one's expression 'I am ϕ -ing/will ϕ ', but does not know what one is doing or will do. So, can we say that, if one knows one's own intentional action, one knows one's own intention? In other words, can we say that one's knowledge of one's intention is transparent to one's knowledge of one's own intentional action? The latter is a question that will be addressed in the next chapter.

Here, we want to return to the PTA and see what else we can add to Boyle's account. Based on PTA, one cannot do something intentionally without knowing how to do it. Thus, one cannot express "I'm ϕ -ing", and ascribe the intention to ϕ to oneself unless one knows how to do ϕ . For example, when someone *does not know at all* how to write/read, she may still make different meaningful sentences regarding writing/reading, and she may judge propositions that are about these actions. She can rightly judge that "X is writing something," "Y is not reading," or "writing is" However, these abilities are not sufficient to self-ascribe the intention to write/read and to express "I'm writing/reading" or "I'll write/read."

However, maybe someone does not know how to write but expresses her intention

to write by claiming that "I'm writing" or "I'll write." To explain why one cannot do something intentionally without knowing how to do it, we have to take into consideration the distinctions between having a desire to do ϕ , trying to do ϕ , and doing ϕ intentionally. As we mentioned earlier in 3.6., contrary to desires, intending to do something requires a commitment to do that action. One can have a desire to ϕ in spite of whether one knows how to ϕ , since having a desire to do something does not bring any commitment to doing that action. However, intending to ϕ requires a commitment to ϕ and doing something in order to do ϕ . The one who does not know how to write cannot give the commitment that will write. Therefore, when there is no commitment, one's claim that "I'm writing" or "I'll write" are more like expressions of one's desires than intentions.³¹

Furthermore, intending to ϕ is also different from trying to ϕ , and we cannot say that someone who is trying to ϕ , is ϕ -ing intentionally. As we explained in 2.5.1.5., when someone knows how to do something, it is meaningless for her to try to do it, instead, she would simply intend to do it. Consider once again our imagined illiterate who, this time, is confident that she is able to write her name, and she does it. Let's begin by explaining with what she means by "I'm able." Given that she does not know how to write, by avowing that she is able, in the best situation, she would mean that she possesses all the circumstantial factors necessary for writing, like having a pen and required physical abilities, and she willing to learn how to write her name. In such a case, normally she does not start out with intending to write her name, but rather she starts out with trying to write it and possibly learning how to do it, then she intends to write. In other words, she believes she can write her name because she thinks it is possible to learn how to write it by trying to do it. When our imagined illiterate does not know how to write her name, she can try to do it by, say, imitating her name written on her mobile bill. Inevitably, once she tries to do it, she is doing some intentional action (for sure, some basic intentional actions), like moving her hand, and so on. However, as long as she is trying to write, she cannot yet say, "I'm writing my name,"³² because she is not confident that what she is doing is really

³¹ See also (Anscombe 1957: 67-8).

³² Consider someone who has a good reason for ϕ -ing. In order to ϕ , one needs to do some other actions that are less complicated than ϕ , let us call them X and Y. Thus, if an agent does X and Y, she consequently is doing ϕ . Imagine that X and Y can only be replaced by either α and β , or γ and δ , which are particular actions. Now, the question is when would we (from a third-person perspective) say that she is trying to ϕ , and when would we say that she is doing ϕ ? From a third-person perspective, we can confidently say that she is ϕ -ing when we observe that she is ϕ -ing, i.e., when we are seeing that she is doing α and β , or γ and δ . Moreover, we usually say that she is trying to ϕ when we see that she is doing some of the actions which would be ended in ϕ -ing, but she does not show sufficient confidence

writing the name.

To elaborate on this point more, suppose that our imagined illiterate wrote her name by chance when she tried it. However, if she did not know how to write and read at all, until the last seconds when she saw the result of her action or somebody else let her know that she did it, she was not confident that she did it. In other words, our imaginary illiterate found out that she was writing her name only through observation, therefore, her action was not intentional.³³ Although when she was trying to write the name, her action under some descriptions was intentional— like drawing some lines on the paper— it was not intentional under this description that 'she was writing her name.'

In short, it seems that until our imaginary illiterate does not know how to write her name, she cannot go further beyond a desire to write it, or trying to write it. Self-ascription of intention to write is possible for her only when she knows how to write. This provides some explanation as to why a satisfying account of self-knowledge of intention must take into account knowledge-how:

▪ **The Minimal Condition on Knowing One's own Intention** (MCKI**):** When one expresses that one is ϕ -ing, one must understand what doing ϕ means and can identify oneself as the person who is doing ϕ , in the sense that one knows how to do ϕ and can identify oneself as the agent who is exercising this knowledge.

But what if our imagined unlettered person who did not know how to write her name, claims that 'I'm writing my name,' and writes her name on the first try? Should we say that she did it unintentionally? This case is so akin to Davidson's carbon copier example that we mentioned earlier. As we said, Davidson brings an example in which a man writing and making ten carbon copies, doing both intentionally, and yet not knowing that he is making copies, but just that he is trying to make them. We have already explained this challenge: Davidson's carbon copier knows how to make one, two, or even nine carbon copies. But he is not sure how to make ten copies. He intends to do it, and he succeed. However, he did not succeed in making ten carbon copies out of the blue. Rather, he expected that if he did again whatever he had done for nine

to be successful. Accordingly, from the first-person perspective, she would say that she is trying to ϕ when there is no confidence that what she is doing will be ended to ϕ .

Consider that, by chance (replacing X and Y, with a bunch of actions), she finally ϕ -ed, by doing α and β . Did she ϕ intentionally? From the first-person perspective, she would find out that she ϕ -ed only by observation; therefore, her action ϕ was not intentional. For one second before succeeding in doing ϕ , she was not confident at all that doing α and β will end in doing ϕ .

³³ See (Anscombe 1957: 13-4).

copies, maybe by a little bit more pursuer on the papers, he would make this time ten copies. We cannot say that Davidson's carbon copier does not know how to do this at all, although a person who makes ten carbon copies successfully every day is more competent in doing this. In other words, knowledge-how comes in degrees.

In the case of our illiterate person, we should take into account some partial/basic knowledge-how for her.³⁴ For example, she knows that if she carefully copies what is written on the bill, she will write her name, and so on. Without considering this partial/basic knowledge of how to write her name, we cannot imagine that she is writing her name. Hence, we need to add a qualification to MCKI* to also cover cases like carbon copier, and reformulate it as below:

▪ **The Minimal Condition on Knowing One's own Intention** (MCKI**):** When one expresses that one is ϕ -ing, one must understand what doing ϕ means and can identify oneself as the person who is doing ϕ , in the sense that one knows how to ϕ , or knows how to do other actions that one might do in order to ϕ , and can identify oneself as the agent who is exercising this knowledge.

To conclude this section, we can say that there are at least two possible solutions for Boyle:

1. Boyle can rule out intentions from his account of self-knowledge
2. Boyle can accept that the condition of understanding in the case of intentions is not understanding but exercising knowledge-how.

³⁴ There are differences between an illiterate person and a talking Parrot. One difference is that the illiterate person has the ability to omit to talk about writing, but the parrot cannot. Parrot, who is taught to say 'He is writing' when it sees somebody is writing, cannot omit to express this sentence. The parrot expresses this passively, and it is just a reaction to what is happening in its environment. However, the illiterate person is the author of the sentence 'He is writing', in the sense that she is expressing it actively based on some evidence she has (of course, other than knowing how to write/read). Furthermore, illiterate people can revise her expression in a way that shows her responsibility in expressing whatever she says.

4.4. Byrne

4.4.1. Introduction

In this section, we will explore Alex Byrne's Transparency account as a means to understand self-knowledge of intention, while also examining its limitations when applied to intentions. As we mentioned in the first chapter, Byrne's transparency account of self-knowledge is an attempt to explain how we have privileged and peculiar access to our own mental states. Byrne argues that the central problem of self-knowledge is to explain how we have this access, and he proposes that the Transparency account can solve this problem.

As we outlined there, Byrne's Transparency account proposes that we can gain knowledge about our own mental attitudes by making inferences from a premise about something in the external world. With regard to our belief, Byrne introduces a specific inference schema, the Doxastic Schema, which allows us to conclude that we possess a belief based on a premise about something in the external world. The Doxastic Schema can be represented as follows:

$$\frac{p}{\text{I believe that } p}$$

The Doxastic Schema suggests that if we discover that 'p', then we can infer that 'I believe that p.' Byrne explains this transition from 'P' to 'I believe that p' in terms of an epistemic rule:

BEL If p, believe that you believe that p.

As elaborated in the first chapter, Byrne believes that the doxastic schema is "self-verifying," which means that even when we attempt to use it to self-ascribe a belief, we end up with a true self-belief, regardless of whether the premise is true or false. This is what explains privileged access, as our beliefs about our own mental attitudes are less fallible compared to beliefs about the minds of others. Peculiar access is explained because the method only works for oneself, as trying to infer that someone else believes p from the premise that p can often be misleading.

While Byrne's transparency account has the potential to explain self-knowledge

of beliefs, it faces limitations when applied to intentions. In the following sections, we will first briefly review Alex Byrne's version of the transparency account of intention, and then we will try to show that his account does not furnish a proper epistemology for intention.

4.4.2. Byrne's Transparency account of Intention

As discussed earlier, Byrne's Transparency account suggests that one can use an inference schema to infer the presence of a particular mental state based on a premise about something in the external world. Byrne argues that different mental states require different schemas. To explain how we come to know about our own intentions, he introduces the Bouletic Schema (BoS). The schema takes its name from the Greek word "Boulomai," meaning 'to will.' This schema can be represented as follows:

$$\frac{\text{I will } \phi}{\text{I intend to } \phi}$$

BoS suggests that if we discover that 'I will ϕ ', then we can infer that 'I intend to ϕ .' Byrne explains this transition from 'I will ϕ ' to 'I intend to ϕ ' in terms of the following epistemic rule:

INT If you will ϕ , believe you intend to ϕ .

According to Byrne, there are two problems of defeasibility³⁵ when it comes to INT. The first problem is that the statement 'I will ϕ ' can be ambiguous, as it may be unclear whether one is expressing an intention or simply making a prediction about what will happen. Byrne uses Anscombe's example to illustrate this ambiguity: "[I]f I say 'I am going to fail this exam' and someone says 'Surely you aren't as bad at the subject as that', I may make my meaning clear by explaining that I was expressing an intention, not giving an estimate of my chances". (Anscombe 1957: 1-2). If 'I will ϕ ' is an instance of prediction, then it cannot provide us with any knowledge about intention.

The second problem of defeasibility concerns cases where an action is a foreseen unintended consequence of another intended action. A person may know that she will

³⁵ What he means by this issue is that "additional evidence (or apparent evidence) can block the inference "from the premise 'I will ϕ ' to the conclusion 'I intend to ϕ .' (Byrne 2011: 215).

carry out ϕ , without explicitly intending to do so, if ϕ occurs as a result of ψ and the individual intends to perform ψ . We have already encountered examples of these actions in section 2.5.2., such as the marathon runner who did not intend to wear out her sneakers or the bomber pilot who did not intend to kill children. In such scenarios, one may believe that one will carry out ϕ , but not believe that one intends to do so.

Based on the examples of the two problems concerning the Bouletic Schema, Byrne concludes that these problems arise when individuals have evidence-based knowledge about their future actions but do not self-attribute the corresponding intention. For example, in the case of the exam failure, the marathon runner, and the tactical bomber, these individuals possess knowledge about their actions based on evidence, without explicitly intending to perform those actions.

Accordingly, Byrne formulates a condition under which the Bouletic Schema is defeasible: "One will not reason in accord with the Bouletic schema if one believes that one's belief that one will ϕ rests on good evidence that one will ϕ ." (Byrne 2011: 218). However, he argues that this defeating condition can be skipped since, as Anscombe notes regarding knowledge of intentional action, one's knowledge of one's own intentional action must not be based on evidence.

Furthermore, Byrne argues that similar to the doxastic schema, the Bouletic Schema provides a plausible explanation for both privileged and peculiar access one has to one's own intentions. He explains that privileged access stems from the strong self-verifying nature of the Bouletic Schema. If one reasons according to the schema while keeping in mind potential defeating conditions (like having good evidence for believing what will be done), then the resulting belief about one's intention will be true.

According to him, peculiar access is explained by the fact that the method only applies to oneself. This is because, in the case of others, the defeating condition is usually present - if we believe someone else will do something, we assume our belief is based on good evidence. Additionally, Byrne argues that inferring that someone else intends to do something is often unwarranted, as she may not have that intention at all. For instance, if someone is going to step on an ant, it is most unlikely that she intends to do so. (Byrne 2011: 219).

In summary, Byrne proposes that one can infer 'I intend to ϕ ' from 'I will ϕ ,' unless one believes that one will ϕ based on good evidence. However, there are criticisms of his view that will be discussed in the following sections.

4.4.3. Objections

One issue with Byrne's Bouletic Schema arises in cases of intentional omissions. The Bouletic Schema struggles to account for the knowledge of our intentions when we deliberately decide not to do something. It appears that the schema falls short in situations where one is intentionally committed to inaction. Our knowledge of what we will do should also encompass our knowledge of what we will choose not to do if we regard omission as a form of action. While some cases of omission can be expressed in terms of performing a specific action, such as choosing to sit down instead of standing up, most omissions cannot be reduced to positive actions— omissions are often the absence of actions.³⁶ Even when one intends to forgo standing up, one's intention is not necessarily to sit in a specific place or maintain a particular position or state, but rather to prevent the action of standing up from occurring in the world. Therefore, the statement 'I will ϕ ' is not suitable to express one's intention when it comes to intentional omissions.

Another issue with this approach is its inability to account for 'intention in action' and 'intention with which.' Byrne refers to Anscombe, stating, "As Anscombe points out, one expresses the intention to ϕ by asserting that one will ϕ ." (Byrne 2011: 215). However, as we discussed earlier, according to Anscombe, one can also express one's intention through an *intentional action/intention in action* ('I'm ϕ -ing') or by the *intention with which* an action is performed ('I'm ϕ -ing in order to ψ '). In this regard, there are particular challenges associated with the use of the future tense, as it cannot express one's belief about what one is currently doing at the moment. If this is the case, then the Bouletic Schema fails to provide knowledge about one's intention with which and intention in action. It appears that the schema only accounts for 'intentions for the future' - those intentions that one has for performing an action in the future.

A further criticism of the Bouletic Schema concerns its reliance on INT. The premise of the schema, as formulated by Byrne, seems to be based on observational evidence and cannot be obtained non-observationally as Anscombe had suggested. As we said earlier, in Byrne's transparency account, generally, the premises of inference schemas are facts about the world, and conclusions are inferred from them according to an epistemic rule. Even if we say that this general approach works with

³⁶ See (Clarke 2010), (Clarke 2012a), and (Clarke 2012b)

regard to other mental states, like beliefs, it does not seem to have application regarding intentions. In the case of the Bouletic Schema, it appears that the INT premise neglects the agential role in executing intentional actions, and the premise can be arrived at through a third-personal perspective, i.e., based on evidence. This means that if the premise 'I will ϕ ' stems from one's knowledge of action, it may not necessarily be an instance of practical knowledge and could be theoretical instead.

The challenge that Byrne faces cannot be overcome simply by stating that 'I will ϕ ' should not be based on evidence. The difficulty lies in the nature of INT, which treats the premise 'I will ϕ ' as an instance of knowledge derived from the world. In short, the intention one self-ascribes based on the Bouletic Schema is an alienated intention, disconnected from one's actual actions— it violates the rules of the direction of fit, which in intention is world-to-mind. To avoid the problem of alienation, the Bouletic Schema needs to incorporate not just a proposition about the world, but also the agent's guiding role in carrying out the action.

In a similar line of criticism, Matthew Boyle (2019) contends that if one wants to self-ascribe an intention to oneself based on the Bouletic Schema, the 'will' in "I will ϕ " should not be a neutral will that merely describes what will happen in the world based on one's assumption that the action will take place. According to Boyle, the premise of the Bouletic Schema already implies the conclusion that is meant to be reached through the inference.

Let's delve into Boyle's objection in greater detail. He differentiates between two senses of "will": an intention-based and a neutral-based sense of will. (Boyle 2019: 12). He asserts that to be justified in the transition from the premise 'I will ϕ ' to the conclusion 'I intend to ϕ ,' one's move should rely on a unique 'intention-based' sense of 'will' used in the premise. Consequently, he believes that "the cognitive transitions we make are not, in fact, transitions from sheer propositions about the world to propositions about the subject's own mental states." This is because if the self-ascription of the intention to ϕ is based on the belief that 'I will ϕ ,' in which 'will' is intention-based, then "although my basis is superficially neutral, it is not genuinely neutral: it presupposes an implicit awareness on my part of what I intend to do" (Boyle 2019: 12).

Given the two last objections, we can attempt to formulate a revised version of the Bouletic Schema. According to the latter objections, in the inference that follows INT, the premise of the schema should not be achieved through theoretical reasoning, and it must be an instance of practical knowledge. Now, we can reformulate INT in this

way: If you have practical knowledge about your action ϕ , believe that you intend to ϕ . As explained in section 3.5., one has practical knowledge when exercising knowledge-how. Thus, we can form the inference in a way that one can self-ascribe an intention to ϕ from the premise that one is exercising one's knowledge-how-to- ϕ . We can formulate INT in the following form:

INT* If you are exercising your knowledge-how-to- ϕ , believe that you intend to ϕ .

INT* is an epistemic rule based on PTA that allows us to move from 'I know that I am ϕ -ing/will ϕ ' to 'I intend to ϕ ' using the Bouletic Schema.

INT* encompasses one's intentions for intentional omissions. One knows that one is intentionally omitting to ϕ , since one is exercising one's knowledge-how-to-omit- ϕ . We will talk more about this point in the next chapter, section 5.5.3.. Moreover, INT* also covers all the descriptions that could be attributed to the intention: Intention with which, intention for the future, and intention in action. We will discuss this point more in the next chapter, section 5.5.4.. Finally, INT* does not violate the rules of the direction of fit in intention because it does not conclude that one has an intention just based on a premise that is a sheer proposition about the world.

Although INT* is in line with PTA, in the next chapter we propose, on the basis of PTA, a more elaborated transparency rule to explain the self-knowledge of intention.

5. Phronetic Rule of Transparency

The primitive sign of wanting is trying to get.

(Anscombe 1957: 68)

The mistake is to think that the relation of being done in execution of a certain intention, or being done intentionally, is a causal relation between act and intention. We see this to be a mistake if we note that an intention does not have to be a distinct psychological state which exists either prior to or even contemporaneously with the intentional action whose intention it is.

(Anscombe 2005: 95)

Abstract

In this chapter, we present a new version of the Transparency account of intention, based on PTA. We argue that knowledge-how not only plays a critical role in knowing one's own intentional actions but also explains the privileged access one has to one's intentions. Towards this end, we follow Setiya in believing that individuals with specific rational capacities can attain self-knowledge by adhering to a rule of transparency. For self-knowledge of intention, we propose that knowledge-how is the particular rational capacity that one must employ and suggest the following transparency rule:

Phronetic rule of transparency (PRT): If one has the capacity to know how to ϕ , and based on this capacity knows that one is ϕ -ing, at the same time and by the same capacity, one can ascribe the intention to ϕ to oneself.

5.1. Introduction

In this chapter, the aim is to present a new version of the Transparency account of intention based on PTA. The argument is that knowledge-how plays a critical role in gaining knowledge of one's own intentional actions and in explaining the privileged access one has to one's own intention. The central idea is that knowledge of intention is grounded in knowledge of intentional action. Accordingly, given that knowledge-how plays a role in achieving knowledge of intentional action, this chapter tries to show that in providing a proper transparency account of self-knowledge of intention, it is necessary to consider the crucial role of knowledge-how.

The account presented here builds on Setiya's account of Transparency in different ways. In his work *'Knowledge of intention'*, valuable insights are offered into how individuals can acquire self-knowledge of their intentions, making a significant contribution to the Transparency method. According to Setiya, an individual's knowledge of her own intentional action (practical knowledge) does not arise from self-knowledge of intention. Instead, self-knowledge of intention arises from practical knowledge through the transparency procedure. However, he endorses that knowledge of intention is groundless and non-inferential. Hence, his transparency approach, contrary to Byrne's, is not an inferential approach. Moreover, contrary to the reflective approach, Setiya does not agree that to have a mental attitude is already to have tacit knowledge of that mental attitude. (Setiya 2012).

Instead of inference and reflection, Setiya's account of Transparency places rational capacities at the forefront. He argues that an individual with a particular rational capacity can attain self-knowledge by adhering to a transparency rule. In other words, when someone exercises a specific rational capacity, she uses it to form a belief about her own mental state by following a transparency rule. In his account of transparency, both the pertinent rational capacity and transparency rule may differ for various mental states. Nevertheless, despite the different transparency rules, Setiya remains loyal to the basic principle of Transparency, which holds that one gains knowledge of one's own mental states not by attending inward to the contents of one's mind, but by attending outward to the world.

Setiya posits that the rational capacity one employs to form beliefs about one's own intention is the capacity to act for reasons. He supports this claim by citing what he

refers to as Anscombe's principle, which states that "the capacity to act for reasons must be a capacity to know what one is doing." (Setiya 2011: 190). Assuming this principle, Setiya argues that an individual who knows her own intentional action also knows her own intention. In other words, he contends that exercising the capacity to act for reasons not only provides knowledge of intentional action but also yields knowledge of intention by following a rule of transparency.

This chapter critically examines Setiya's attribution of the capacity to act for reasons as the source of groundless knowledge of intention. The argument is that this capacity cannot be the rational capacity responsible for groundless knowledge of intention. Instead, we propose that the capacity to know-how is the true source of groundless knowledge of intention. Furthermore, we will demonstrate how a rule of transparency based on PTA³⁷ can better explain our knowledge of our own intentions.

This chapter consists of five main sections. Section 5.2 depicts Setiya's starting point for linking knowledge of intention and knowledge of intentional action. His aim is to address the claim that knowledge of intentional action is inferred from knowledge of intention. Setiya argues that explaining one's privileged access to one's own intention demonstrates that knowledge of intentional action is not based on an inference from intention. To account for this privileged access, Setiya appeals to Transparency. Section 5.3 explains his approach to Transparency by outlining his transparency account for belief. Section 5.4 describes how Setiya extends the transparency approach from the cognitive to the practical sphere, specifically from belief to intention. In Section 5.5, an alternative rule of transparency for intention is presented, followed by a critical examination of Setiya's view. Five difficulties facing his account are outlined, and the alternative rule of transparency for intention is defended.

³⁷ Phronetic Theory of Action (PTA): S intentionally ϕ -ed at T if and only if S ϕ -ed at T and S (to some degrees) knew how to ϕ and S exercised his/her knowledge-how.

5.2. Non-observational but Inferential

In section 3.1., we mentioned that one of Anscombe's novelties was to demonstrate the proximity between knowledge of intentional action (as knowledge of something that occurs not in mind but in the external world) and self-knowledge (as knowledge of one's own mental states), in the respect that both are non-observational. Furthermore, we talked about the problem of self-knowledge and the problem of knowledge of intentional action. We said that these problems, respectively, concern how one often has a unique way to gain knowledge of one's own mind differently from the third-person way, and the question of how one has a particular way to know one's own intentional actions divergent from the knowledge one has of one's unintentional actions as well as third-person's intentional and unintentional actions. While Anscombe's theory may suggest that a solution to the problem of self-knowledge could be applied to the problem of knowledge of intentional action and vice versa, Setiya disagrees.

In Chapter Three, Setiya's explanation of the possibility of non-observational knowledge in intentional actions was discussed. We showed how he succeeded in demonstrating that practical knowledge is reliable and defensible by relying on knowledge-how and its epistemic function in justifying beliefs in intentional actions. Setiya did not adopt a Transparency approach in explaining practical knowledge. However, he finds the transparency method relevant in addressing the problem of self-knowledge in general and knowledge of intention in particular.

Setiya's main aim is to defend Anscombe's notion that knowledge of intentional action is non-inferential when addressing the issue of self-knowledge of intention. Some philosophers do not perceive any contradiction between being non-observational and inferential. In her paper "How We Know What We're Doing," Sarah Paul offers an inferential theory of non-observational knowledge of intentional action. According to Paul, non-observational knowledge of one's intentional action means that one does not require "to appeal to sense evidence" to have this knowledge— this knowledge is non-perceptual. (Paul 2009: 2). Paul believes that the non-observational nature of knowledge of intentional action does not preclude the possibility that it is obtained through inference. According to Paul, an individual's non-observational knowledge of her intentional action is partially based on her intention in action. The individual is likely to believe that "I'm ϕ -ing," based on an inference from her intention

to ϕ , along with other epistemic factors like knowledge of ability. (Paul 2009: 14). Falvey also holds a similar view, arguing that there is a continuity between intending and doing, and an individual's knowledge that she is ϕ -ing largely relies on her knowledge that she intends to be ϕ -ing. (Falvey 2000).

If we want to formulate the idea that one's knowledge of doing ϕ is primarily derived inferentially from her knowledge of intending to ϕ , our inference would be as follows:

I have the intention to ϕ

I know that I am able to ϕ (or other epistemic factors like this)

I am ϕ -ing/I will ϕ

Setiya, in his paper "Knowledge of Intention," criticises the view that knowledge of intentional action is derived from knowledge of intention. Instead, he argues that knowledge of intention is based on knowing what we are doing or what we are going to do. Setiya uses the Transparency method to explain how knowledge of intention can be obtained from knowledge of intentional action. He contends that "knowledge of intention is transparent to knowledge of action; it is by knowing what we are doing, or what we are going to do, that we know what we intend." (Setiya 2011: 177). Prior to examining how Setiya employs the Transparency method to explain knowledge of intention, we will first explore how he establishes the foundation for his explanation by addressing the transparency of belief.

5.3. The Transparency of Belief

Setiya initiates his explanation of knowledge of intention by addressing how one acquires knowledge of one's own beliefs. He asserts that one's knowledge of one's own beliefs is typically groundless, meaning that it is not based on "quasi-perceptual appearances of belief or on inference from evidence of other kinds." (Setiya 2011: 178). In other words, Setiya claims that one's knowledge of one's own beliefs is often groundless, in the sense that one's belief that one believes that *p* is not based on one's perception of one's belief that *p* or an inference from one's other beliefs. (Setiya 2011: 178). However, how is this groundless knowledge possible?

Having this question in mind leads Setiya to consider Shoemaker's assertion that *self-blindness* is impossible. In his paper 'On Knowing One's own Mind,' Shoemaker defends the Cartesian view that an individual's access to her own mental states is distinct from her access to the mental states of others.³⁸ Shoemaker agrees that an individual has privileged access to her own mental states— *direct* access that he calls *self-acquaintance*. (Shoemaker 1988: 184). In order to support self-acquaintance, one must argue against the possibility of *indirect* access to one's own mental attitudes in a third-person manner, or as Shoemaker phrases it, one must argue against *self-blindness*. To explain what self-blindness is and what it means to be self-blind, Shoemaker describes that a "self-blind creature would be one which has the conception of the various mental states, and can entertain the thought that it has this or that belief, desire, intention, etc., but which is unable to become aware of the truth of such a thought except in a third-person way." (Shoemaker 1988: 189). Setiya finds inspiration in Shoemaker's concept of self-blindness and asserts that if one is capable of reasoning (has the capacity for inference) and has the concept of belief (can ascribe beliefs to others), then one has first-person access to one's own beliefs. (Setiya 2011: 180). He regards this as "a datum in the study of self-knowledge" and does not see a need to argue for it. Based on this datum, Setiya lays down the following principle:

Cognitive Self-Knowledge: "If A has the capacity for inference and can ascribe beliefs to others,³⁹ she has the capacity for groundless knowledge of her own beliefs." (Setiya 2011: 180)

According to Setiya, groundless knowledge of one's own beliefs must emanate

³⁸ We explained this view earlier in 1.1.1.2.

³⁹ Namely, having the relevant concepts.

from the exercise of a capacity that rational beings possess. (Setiya 2011: 181). Following the principle of Cognitive Self-Knowledge, the rational capacity that can form groundless knowledge of belief should be constrained to the capacities one exploits in making inferences and applying the concept of belief. Since the mere concept of belief cannot bring a new way of forming beliefs, he concludes that the rational capacity must come from the capacity for inference. (Setiya 2011: 181). However, he had already defined the groundless knowledge of belief as a sort of non-inferential knowledge. How can inference be the source of this groundless knowledge while this knowledge is not inferential? In Setiya's opinion, it might seem paradoxical that appealing to inference can explain why our knowledge of belief is non-inferential. However, he thinks that Transparency helps us to solve this paradox. His solution has two parts:

"The first is to recognize inference as a species of epistemic rule-following: the application to evidential rules of a more general capacity to form beliefs on the basis of other beliefs. ... The second step is to formulate a rule of transparency for belief inspired by Evans: "whenever you are in a position to assert that *p*, you are ipso facto in a position to assert 'I believe that *p*.'"" (Setiya 2011: 183-4).

We already talked at length about the second step in the first chapter and are familiar with it. But regarding the first step, the difference between epistemic rule-following and inference may not be obvious without further clarification. In 1.4.4., we explained that an epistemic rule is a general principle or guideline that one can follow to acquire knowledge or justify beliefs. Following an epistemic rule could make it possible to form a belief on the basis of another belief. For instance, consider Byrne's Doxastic Schema:

$$\frac{p}{\text{I believe that } p}$$

According to Setiya, we may think of Byrne's doxastic schema not as an instance of inference in its familiar sense that the premisses provide evidence for conclusions, but as an instance of 'epistemic rule-following.' In this sense, we do not "take the 'premiss' of the schema, that *p*, as evidence for its 'conclusion', or ... [as] the ground of [our] self-ascription in any sense other than being the content of the belief from which the self-ascription derives." (Setiya 2012: 266). One can form a belief about one's own belief that '*p*', by following a rule whose premise is "known to be no evidence for its truth." (Ibid.). Accordingly, Setiya explains that there are two main differences

between instances of epistemic rule-following and inference: First, in an inference from the premise that *p* to the conclusion that 'I believe that *p*,' the transition relies on sufficient evidential support. However, in an epistemic rule-following transition, the premise that '*p*' does not need to be evidential support for the conclusion that 'I believe that *p*.' Second, in an inference, whether the conclusion is justified or not depends on whether the premise is justified or not. However, in an instance of epistemic rule-following, since we are following the epistemic rule, the conclusion that 'I believe that *p*' is justified, even if our belief that '*p*' is not. (Setiya 2011: 184). We explained the second feature when we explained the self-verifying nature of Byrne's Doxastic Schema.

After showing the distinction between epistemic rule-following and inference, Setiya can explain the connection between the first step and the second one and clarify the role of the capacity for inference in self-knowledge of belief. If we recognise inference as a species of epistemic rule-following, then one who has the capacity for inference has the capacity for following an epistemic rule. In self-knowledge of belief, the epistemic rule that is followed is the rule of transparency for belief, i.e., "whenever you are in a position to assert that *p*, you are ipso facto in a position to assert 'I believe that *p*.'" In this regard, one who has a belief can ascribe this belief to oneself.

Now that we have seen how Setiya explained groundless knowledge of belief, it is time to see how he explains one's knowledge of one's own intention. In the next section, we will see how Setiya adopts the transparency approach from the cognitive to the practical sphere.

5.4. The Transparency of Intention

Setiya aims to apply the transparency approach to the practical scope by formulating a rule of transparency that enables us to self-ascribe our intentions. He starts by establishing a general principle based on Anscombe's theory, which he calls "Non-Perceptual Knowledge." According to this principle, if an agent has the capacity to act for reasons, then she also has the capacity to know what she is doing without relying on observation. To gain support from critics like Paul and Falvey on this principle, Setiya interprets knowledge without observation as knowledge that does not depend on sufficient perceptual evidence.

"Non-Perceptual Knowledge: If A has the capacity to act for reasons, she has the capacity to know what she is doing without observation—in that her knowledge does not rest on sufficient perceptual evidence." (Setiya 2011: 189).

We previously discussed in detail Anscombe's account of intentional action and knowledge of intentional action in chapters two and three. The principle of *Non-Perceptual Knowledge* is based on Anscombe's theory that what we do for reasons we do intentionally, and what we do intentionally we know without observation.

The principle of *Non-Perceptual Knowledge* is acceptable to critics like Paul and Falvey, who, as we mentioned earlier, present an inferential account of one's knowledge of one's own intentional action. As we noted there, the critics disagree with Setiya regarding the non-inferential character of knowledge of intentional action—they believe that knowledge of action is non-perceptual, but is obtained by inference from the agent's relevant intention. As a result, these critics should concur that one who has the capacity for non-perceptual knowledge of intentional action also has "a capacity for inference from intention". (Setiya 2011: 189). Therefore, assuming the principle of *Non-Perceptual Knowledge*, they should agree that the one who has the capacity to act for reasons has the capacity to make inferences based on intention. Making inferences based on intention demands knowledge of that intention. Consequently, someone with the capacity to act for reasons possesses the capacity for knowledge of intention and can draw inferences from that intention.

Furthermore, based on what we mentioned regarding the impossibility of self-blindness, knowledge of intention, like knowledge of belief, should be groundless. Therefore, if one has the capacity to act for reasons and has the concept of intention, one has first-person access to one's own intentions. In this regard, Setiya lays down

the Practical Self-Knowledge principle:

Practical Self-Knowledge: If A has the capacity to act for reasons and can ascribe intentions to others, she has the capacity for groundless knowledge of her own intentions. (Setiya 2011: 189).

Setiya asserts that groundless knowledge of one's own intention, similar to groundless knowledge of belief, must arise from the exercise of a capacity inherent to rational beings. In line with the principle of Practical Self-Knowledge, the rational capacity responsible for forming groundless knowledge of intention must be restricted to the capacities used in acting for reasons and applying the concept of intention. Since the mere concept of intention cannot introduce "a new way of forming beliefs," the capacity to act for reasons ought to be the source of groundless knowledge of intention (Setiya 2011: 190).

Having identified the rational capacity that serves as the origin of groundless knowledge of intention, Setiya aims to apply the transparency approach from the cognitive domain to the practical realm. This adaptation consists of two steps:

The first step involves embracing Anscombe's Principle, which states that "the capacity to act for reasons must be a capacity to know what one is doing—not just what one intends—without sufficient prior evidence." (Setiya 2011: 190).

The second step entails formulating a rule of transparency for intention, drawing inspiration from Evans' perspective on the self-ascription of perceptual experience. As Evans says:

[A] subject can gain knowledge of his internal informational states in a very simple way: by re-using precisely those skills of conceptualization that he uses to make judgements about the world. Here is how he can do it. He goes through exactly the same procedure as he would go through if he were trying to make a judgement about how it is at this place now, but excluding any knowledge he has of an extraneous kind. (That is, he seeks to determine what he would judge if he did not have such extraneous information.) The result will necessarily be closely correlated with the content of the informational state which he is in at that time. Now he may prefix this result with the operator 'It seems to me as though . . .'. This is a way of producing in himself, and giving expression to, a cognitive state whose content is systematically dependent upon the content of the informational state, and the systematic dependence is a basis for him to claim knowledge of the informational state. (Evans 1982: 227-28)

In essence, Evans proposes that individuals can achieve knowledge of their perceptions by redeploying the capacity that they exploit to make judgments about the world. To adopt this rule for intention, Setiya modifies it. According to him, an

agent can achieve knowledge of her own intention by exploiting the same capacity that she exploits to make judgments about what she is doing. Setiya reformulates Evans' remark as follows:

"He goes through exactly the same procedure as he would go through if he were trying to make a judgement about [what he is doing] but excluding any knowledge he has of an extraneous kind. . . . The result will necessarily be closely correlated with the content of [his intention.] Now he may prefix this result with the operator "I intend . . .". This is a way of producing in himself, and giving expression to, a cognitive state whose content is systematically dependent upon the content of the [intention], and the systematic dependence is a basis for him to claim knowledge of [what he intends]." (Setiya 2011: 193).

The formulation implies that as soon as one knows what one is doing while exercising one's capacity to act for reasons, one can access one's own relevant intention by using the same capacity. According to the rule of transparency for intention, this capacity serves as the source for both one's knowledge of one's own intentional action and one's knowledge of one's own intention. However, the mechanism by which this process occurs remains unclear—is it simply by appending the content of one's knowledge of one's action with the operator "I intend..."? Additionally, how does one initially obtain knowledge of one's own intentional action? Setiya's response to this question is implicitly embedded within his explanation of the rule of transparency for intention.

Setiya elucidates it as follows:

RTI: "...When I employ the capacity to act for reasons in forming an intention and thus acquire some degree of belief—ideally, knowledge—about what I am doing and why, I may at the same time and by the same capacity form the belief that I have that intention" (Setiya 2011: 193-4).

This explanation demonstrates Setiya's perspective on how one can acquire knowledge of one's own intentional action. It is grounded in his theory about one's knowledge of one's own intentional action, as discussed in his papers "Practical Knowledge" and "Practical Knowledge Revisited" and mentioned in section 3.5.. In these works, he posits that the intention to ϕ involves the belief/partial belief that 'I am ϕ -ing' or at least 'I am ψ -ing,' when ψ is done in order to ϕ . Furthermore, to explain how the belief involved in intention amount to knowledge, he maintained that this belief is supported by knowledge-how-to- ϕ or at least knowledge-how-to- ψ when ψ is done in order to ϕ . Based on this view, we can say that Setiya's argument for explaining how one has groundless knowledge of one's own intention proceeds as

follows:

- I. One has a reason for ϕ -ing and forms an intention to ϕ based on one's own reason.
- II. An intention to ϕ involves the belief that 'I am ϕ -ing' or at least 'I am ψ -ing,' when one does ψ in order to ϕ .
- III. This belief is supported by knowledge-how-to- ϕ or at least knowledge-how-to- ψ when one does ψ in order to ϕ .
- IV. One knows that one is ϕ -ing.
- V. The content of one's knowledge that one is ϕ -ing is systematically dependent upon the content of one's intention to ϕ .
- VI. One forms the belief that one has the intention to ϕ .

Through these steps, Setiya appears to have partially succeeded in outlining how one can acquire groundless knowledge of intention. While this study fully agrees with his progression from practical knowledge towards knowledge of intention, it questions the steps he has taken in that direction. If this outline is meant to establish the rule of transparency for intention, it seems that some details are still missing. There are five questions that can be raised regarding Setiya's account:

1. What exactly is the capacity to act for reasons? How can one exercise this capacity, and what does exercising this capacity bring about?
2. How can one gain knowledge of ϕ -ing when one does not have even partial belief that 'I am ϕ -ing,' and the only belief that one has is that 'I am ψ -ing,' when one is ψ -ing in order to ϕ ?
3. How can the capacity to act for reasons be the source of one's knowledge of one's own intention in cases where one intentionally omits to do something?
4. How can one know one's own prospective intentions without starting to perform any action?
5. How does the transition from practical knowledge to knowledge of intention occur?

This study believes that the key issue with Setiya's account lies in his emphasis on the capacity to act for reasons as the source of groundless knowledge of intention. In the following discussion, we will first propose an alternative rule of transparency for intention based on PTA, suggesting that the capacity to know-how could also be considered as the source of self-knowledge of intention. Then, we will critically examine Setiya's view by addressing these five questions. Concurrently, we will

defend the alternative rule of transparency for intention, demonstrating how exercising knowledge-how can be the true source of groundless knowledge of intention and why a rule of transparency based on PTA provides a better explanation for our access to our own intentions.

5.5. Phronetic rule of transparency

So far, this dissertation concurs with Setiya that knowledge of intentional action is not derived from an inference based on knowledge of intention; instead, knowledge of intention is contingent on knowledge of intentional action. Additionally, this study recognises that knowledge of intention is groundless: since individuals have first-person access to their own mental states and self-blindness is impossible, one's knowledge of one's own intention is neither perceptual nor inferential. Like Setiya, this dissertation also posits that a rational capacity could serve as the source of both knowledge of intention and knowledge of intentional action. However, this dissertation challenges Setiya's assertion regarding the role of the capacity to act for reasons in acquiring knowledge of intention. This study disagrees with him that this capacity could be the very rational capacity sought as the source of groundless knowledge of intention. In the following subsections, by demonstrating its inefficiency and inadequacy, we will explain why the capacity to act for reasons is not the best choice for the role it has been given. Alternatively, we consider the capacity to know how to do something fit for this role.

In the second chapter, we explained that based on PTA, exercising knowledge-how-to- ϕ entails doing ϕ intentionally—S intentionally ϕ -ed at T if and only if S ϕ -ed at T, S (to some degree) knew how to ϕ , and S exercised her knowledge-how. Moreover, we said that intentionally doing ϕ involves the belief/partial belief that 'I am ϕ -ing.' It is not necessary that, as proponents of the doxastic account of intention, we consider having this belief dependent on having the intention to ϕ . Whether intention involves the belief or not is not relevant here. What is relevant is that without believing, to some extent, that 'I am ϕ -ing,' it is impossible to ϕ intentionally. Finally, we mentioned that this belief is supported by knowledge-how/partial knowledge-how-to- ϕ . This was the way PTA explained knowledge in intentional action.

Furthermore, Anscombe does not merely claim that doing something intentionally is doing it for a reason. She also asserts that doing something intentionally is doing something knowingly, and she acknowledges that in doing something intentionally, one exercises one's own knowledge-how. According to Anscombe, "a man has practical knowledge who knows how to do things; ...; but if we hear of a capacity, it is reasonable to ask what constitutes an exercise of it. ... In the case of practical knowledge, the exercise of the capacity [to know-how] is nothing but the doing or

supervising of the operations of which a man has practical knowledge; ..." (Anscombe 1957: 88). Consequently, we can assert that even critics who believe that non-perceptual knowledge of intentional action inferentially follows from a premise about one's intention should not have a problem with the reformulation of Non-Perceptual Knowledge as follows:

Non-Perceptual Knowledge': If A has the capacity to know-how, she has the capacity to know what she is doing without observation— in that her knowledge does not rest on sufficient perceptual evidence.

Now, if critics maintain that non-perceptual knowledge of action is based on an inference from the agent's relevant intention, then it should be acceptable for them to say that one who has the capacity to know-how has the capacity to make inferences based on intention. Making inferences based on intention demands knowledge of that intention. Consequently, someone with the capacity to know-how possesses the capacity for knowledge of intention and can draw inferences from that intention. Furthermore, since self-blindness is impossible, one's knowledge of one's own intention is groundless. Therefore, if one has the capacity to know-how and has the concept of intention, one has first-person access to one's own intentions. Now, we can also reformulate Practical Self-Knowledge as below:

Practical Self-Knowledge': If A has the capacity to know-how and can ascribe intentions to others, she has the capacity for groundless knowledge of her own intentions.

Now, if we suppose that groundless knowledge of one's own intention must emanate from the exercise of a capacity that rational beings possess, the capacity to know-how could be the very capacity. Then, for adopting the transparency approach toward the practical sphere, we simply need to consider two steps: The first step is accepting that the capacity to know-how must be a capacity to know what one is doing— not just what one intends— without sufficient prior evidence. The second step is to follow the same rule of transparency which Setiya borrowed from Evans. The outcome can be demonstrated by revising RTI based on PTA:

Phronetic rule of transparency (PRT): If one has the capacity to know how to ϕ , and based on this capacity knows that one is ϕ -ing, at the same time and by the same capacity, one can ascribe the intention to ϕ to oneself.

We can sketch out the procedure suggested by *PRT* as follows:

- I. One exercises one's knowledge-how-to- ϕ .
- II. One is doing ϕ intentionally.

- III. One has this belief that 'I am ϕ -ing.'
- IV. One's belief is supported by one's knowledge-how-to- ϕ .
- V. One knows that one is ϕ -ing.
- VI. One forms the belief that one has the intention to ϕ .

Now, the question is whether appealing to PRT can lead to a better explanation of self-knowledge of intention. In what follows, we will answer this question by returning to the five questions challenging RTI.

5.5.1. In Search of a Transparent Rational Capacity

One of the greatest challenges facing Setiya's view is figuring out how the capacity to act for reasons can launch the procedure of gaining groundless knowledge of intentional action and intention. It is unclear what it means to have this capacity and what capabilities a person must have to attribute it to herself. According to Anscombe's remarks, we can say that 'what we do for reasons, we do intentionally.' Therefore, having the capacity to act for reasons is having a capacity to perform intentional actions. However, without further explanation, it is not instructive to say that a capacity to perform intentional actions is the source of our groundless knowledge of intentional actions and intentions. Setiya needs to explain what exercising the capacity to act for reasons is, what exercising this capacity brings about, and how it can be the source of groundless knowledge.

There are some points about this capacity that we can conjecture. Obviously, having this capacity is not enough for doing something intentionally, and one should also exercise this capacity. One would have the capacity to act for reasons but does an action unintentionally, while the exercise of this capacity brings about an intentional action. Furthermore, since this capacity is also the source of the groundless knowledge of intentional action, exercising it should also entail this knowledge. Therefore, exercising the capacity to act for reasons should entail both an intentional action and knowledge of it.

Consequently, it also becomes clear that employing the capacity to act for reasons is not identical to having reasons for doing something, since the latter does not necessarily form an intention nor entail an intentional action. This dissertation agrees

with the idea that whatever we do intentionally, we have a reason for doing it.⁴⁰ But the reverse is not true. Someone could have reasons for ϕ -ing while she still does not intend to ϕ . Consider someone who has good reason to quit smoking. However, this person never forms an intention to quit cigarettes and does not do anything toward it. Therefore, employing the capacity to act for reasons should be more than just having reasons to act.

Furthermore, the capacity to act for reasons consists of at least two different capacities, a non-rational and a rational capacity. One who can act for reasons should be able to act according to one's reasons. A paralysed person who has good reason for saving herself from a terrible situation, like a building that is catching on fire, cannot act based on her reasons. Furthermore, to act for reasons, one should know how to do the action and exercise it. If someone does not know how to escape from a building that is catching on fire, she does not have the capacity to escape, as well. As long as it is not specified that the capacity to act for reasons is anything beyond the capacities it comprises, we can delegate the role that Setiya has assigned to the capacity to act for reasons to the other capacities that constitute this capacity.

Furthermore, as we mentioned in the second chapter, Setiya is aware of the important epistemic role that the ability to do something and knowing how to do it play concerning knowledge of intentional actions. However, when he is talking about the capacity to act for reasons, he considers *the ability to do something* and *knowing how to do it* untied:

“Epistemic right to form the belief that I am doing ϕ by forming an intention turns on knowing how to ϕ , not just the general capacity to act for reasons. And even when I do know how, this right may be revoked or undermined: I may have doubts about my own ability or evidence of interfering factors.” (Setiya 2011: 192).

Although Setiya disagrees that having the capacity to act for reasons necessarily involves having the pertinent ability and knowledge-how, he emphasises that the capacity to act for reasons is not all one needs to judge what one is doing. If this capacity alone is not the source⁴¹ of our knowledge of what we intentionally do, and knowledge of intention depends on knowledge of intentional action, a natural question arising here is why should it not be said that all the capacities involved in

⁴⁰ As Anscombe mentioned, there are cases that one may respond to the “why?” question something like ‘I had no particular reason.’ But it is not relevant here.

⁴¹ Setiya does not say that it is a source but he says “the capacity to act for reasons **must be a capacity** to know what one is doing— not just what one intends— without sufficient prior evidence.” (Setiya 2011: 190).

this knowledge, together are the source of our knowledge of intention? This question is irrelevant only if there is a capacity that its exercise alone can entail intentional action and knowledge of it. In other words, if the capacity to act for reasons alone cannot be the source of our knowledge of what we intentionally do, it makes more sense that we consider another capacity to be the source of groundless knowledge of intention as long as exercising the other capacity entails both intentional action and knowledge of it solely.

We said that the capacity to act for reasons consists of being able to do something and knowing how to do it, and as long as we do not have a clearer picture of its nature, we can identify it with these two capacities. Now, if the epistemic right to form the belief that 'I am doing ϕ ' also turns on these two capacities, we should investigate whether exercising one of them alone can entail both an intentional action and knowledge of it. In this way, we can see whether one of these capacities can be the source of knowledge of intention.

Being able to do something cannot be the source of knowledge of intention since it is not a rational capacity. Therefore, we just need to see whether exercising knowledge-how can solely be the source of one's knowledge of one's own intentional action. In the second and third chapters, we talked about the relationship between having the ability to do something and knowing how to do it. We said that one might be able to do something but does not know how to do it, even partially. However, based on PTA, one's action would be unintentional in such cases. Also, we said that it is impossible to attribute knowledge-how-to- ϕ to oneself and not be able to ϕ . This is because, in such a case, the required knowledge is how to ϕ with the disability, whereas the knowledge that here one might attribute to oneself is knowledge-how-to- ϕ if the circumstances were otherwise and one were able to ϕ . Moreover, in cases where someone is entirely unable to ϕ , she cannot even form an intention to ϕ , and a desire to ϕ should be the appropriate mental state that she can ascribe to herself.

In summary, we can conclude that having knowledge-how implies knowledge of ability because it is knowledge-how to do the thing in the current circumstances, whereas just being able to do something not only does not guarantee knowing how to do it but also cannot lead to doing it intentionally without exercising knowledge-how.⁴² Hence, we can overlook the epistemic role of knowledge of ability in forming

⁴² Here it is evident that we refrain from including knowledge-how in having the ability to do something. By having the ability, we mean having all other necessary conditions to do the thing. Otherwise, if we consider having pertinent knowledge-how as a condition for being able to do something, a lack of this knowledge indicates a lack of ability.

the belief that 'I am doing ϕ ' in favour of knowledge-how.

In the previous chapters, we explained how exercising knowledge-how entails both an intentional action and groundless knowledge of it. Consequently, as long as exercising the capacity to know how to do something solely entails both intentional action and knowledge of it, it makes more sense to consider this capacity the source of groundless knowledge of intention. We can also say that in contrast to the capacity to act for reasons, the capacity to know how to ϕ is less ambiguous in nature.

Nevertheless, we still need to look at the other problems facing RTI and see if PRT can overcome them. In the following subsection, we will demonstrate the shortcomings of RTI in explaining one's knowledge of one's own intentional action.

5.5.2. Practical Knowledge is not Reducible

The second issue that RTI is facing is the complexity that we can find in Setiya's explanation regarding one's knowledge of one's own intentional action: how can one gain knowledge of ϕ -ing when one does not have even partial belief that 'I am ϕ -ing,' and the only belief that one has is that 'I am ψ -ing,' when one is ψ -ing in order to ϕ ? This difficulty is also mirrored when instead of knowing how to ϕ , the agent only knows how to ψ . Knowing how to ψ does not support the belief that 'I am ϕ -ing' unless the agent partially knows how to ϕ and, to some degree, believes that ψ -ing will end in ϕ -ing. We talked about it in the third chapter. In 3.6., we mentioned that the object of the agent's (partial) belief, when she is ϕ -ing intentionally, must be the very 'I am ϕ -ing' (not 'I am ψ -ing'), and what supports this belief is the very agent's (partial) knowledge-how-to- ϕ (not the agent's knowledge-how-to- ψ).

PRT is not confronted with the second difficulty that Setiya's rule of transparency is facing. For, PRT does not need to reduce the belief that 'I am ϕ -ing' or knowledge-how-to- ϕ , to the belief that 'I am ψ -ing' or knowledge-how-to- ψ , when one is ψ -ing in order to ϕ . We explained in the third chapter that Setiya does not need to provide a weaker reading of his thesis on knowledge of action. Anscombe's insight was that our knowledge of intentional action is not based on observation. However, it is essential to note that this kind of knowledge is specific to the very action we are currently doing intentionally, and cannot be reduced to knowledge of other intentional actions we may be doing in order to achieve the intended result. In other words, our knowledge

of intentional action ϕ cannot be reduced to our knowledge of intentional actions ψ , even if we are doing ψ in order to accomplish ϕ .

In the following subsection, we address another challenge facing RTI regarding intentions in omission. We will show that omissions are not actions, and Setiya's view cannot cover knowledge of intentions in intentional omissions. Furthermore, we will explain how PRT can overcome this challenge by appealing to the agentic role of the subject.

5.5.3. Intention in Omission

The third issue that RTI is facing concerns intention in omission. Setiya makes no attempt to account for knowledge of intentions that one has in intentional omissions. One may intentionally commit to doing nothing and have a groundless knowledge of one's intention. In this regard, Setiya's account of groundless knowledge of intention would not cover intentions in omissions unless we consider omissions as actions. If omissions are not actions, then instead of the capacity to act for reasons, Setiya should appeal to something like *the capacity to not act for reasons*. In addition, since omissions are not actions, he would also need to explain one's knowledge of one's own intentional omission.

Since the definition of the term 'omission' varies among philosophers, it is important to clarify how the term is used here. Some philosophers, like Vermazen (Vermazen 1985), hold 'omissions' as cases of negative action. In the literature, the term negative act tends to be used to refer to what an agent does not do, like not-sitting, not-walking, or not-raising a hand. Since an agent has not done an action, some philosophers want to say the agent has done something of the negative kind. For example, Vermazen says: "[I]f I intentionally pass up a chance to win at cards by laying down the ten of clubs, I have done something — performed an act— describable as not laying down the ten of clubs and as not bringing about my winning." (Vermazen 1985: 93-4). For Vermazen, not laying down the ten of clubs and not bringing about his winning are the negative acts he has done by omitting to win at cards by laying down the ten of clubs.

Douglas N. Walton provides a similar explanation concerning omission. He says, "if I omit to do a, this means that I do not bring it about that p, where p describes what

is brought about in a. Thus an omission is a not-doing.” (Walton 1980: 325). However, as Randolph Clarke disputes, there is no need to take *not bringing it about that something* to be a kind of action. He says, “one thing’s not causing another isn’t a kind of negative causing; it isn’t a kind of causing at all.” (Clarke 2012b: 129). If we consider that actions are changes in the world, no changes are brought out in the world by omitting to do something.

However, some philosophers believe that, by omitting, one brings about changes in the world. According to this group of philosophers, omissions are the very positive actions that one then performs. Positive actions can be defined as everyday actions that are usually addressed in action theory and are the centre of attention in the literature, like raising a hand, walking, or sitting. For the proponents of this approach, if someone intentionally passes up a chance to win at cards by not laying down the ten of clubs, she has done something — performed an act— describable as, for example, laying down the nine of clubs or letting the opponent win.

Nevertheless, in most cases of omission, we cannot regard what is intended as identical to positive action. Intentionally omitting is omitting for reasons⁴³, and as Clarke mentions, “in some cases the reasons for which one performs a certain action are different and independent from the reasons for which one omits to perform (or refrains from or forbears performing) a different one.” (Clarke 2012b: 136) The reason one would have to lay down the nine of clubs might be different from the reason one has had for omitting to lay down the ten. For example, perhaps the agent omits to lay down the ten of the clubs to pass up a chance to win at cards, but by laying down the nine, she wants to send a signal to her friend and makes her friend aware of her decision. Furthermore, If the card player intends to omit laying down the ten of clubs, she simply needs not lay down any card. She does not need to play another card unless she has another purpose.

Showing the other flaws of these approaches can provide us with a better grasp of intentional omissions, but it is beyond the scope of what we are going to do here. In short, this study agrees that there are some cases of intentional omission that one could express in terms of doing something, positive or negative; for instance, when one omits to stand up by sitting on a chair or omits to walk by staying still. However, omissions are not identical to actions, since there are cases of intentional omission, that we cannot even express them in terms of positive or negative actions, and

⁴³ See (Clarke 2010: 171), (Clarke 2012a: 362), and (Clarke 2012b: 128)

certainly do not fall under any of these approaches.

In this regard, we can say that omissions are neither negative actions nor positive ones. As a result, the capacity to act for reasons cannot be the source of our groundless knowledge of our intention to omit, and Setiya needs to modify his account in a way that covers this kind of intention. We already mentioned that one solution for Setiya could be appealing to a capacity like *the capacity to not act for reasons*. If that is the case, depending on whether the agent acts intentionally or omits intentionally, one of the two capacities is the source of groundless intention. However, knowledge of intention cannot have two different sources, and we need to see what is in common between these two capacities. Thus, to find out what is in common, it is rational to ask about the common trait of intentional actions and intentional omissions. This is why we need to define omissions and say if omissions are not actions, what omissions are.

According to Clarke (2012), an omission is “an absence of an action,” so that “there’s nothing in the world that is the omission.” (Clarke 2012a: 361). This definition is consistent with saying that the agent sometimes does some actions in omitting to do something. What is absent in omissions is the very action that has been omitted. In other words, it can be said that omitting to perform an action is sometimes accompanied by performing some present actions, but it cannot be said that performing such actions makes what is omitted present in the world. For instance, to omit to wake up at seven in the morning, an individual might turn off the phone alarm. Although turning off the alarm is something present, she does it in order to keep waking up in the early morning absent. In the same way, we can say that omitting to act is sometimes accompanied by performing some other omissions. Imagine that the individual always sets her alarm clock for seven. But now she intends to omit waking up at that time and intend not to set the alarm tonight.

Intentionally doing or not doing something in order to omit doing something else reveals the agentive role of the subject. Analogous to intentional actions in which the agent takes the necessary steps toward performing the action or supervises the operations she is doing or going to do, in a similar way, when she intentionally omits, she takes the necessary steps toward omitting the action or supervise operations in a way to keep a particular action absent. Consequently, concerning intentional omissions, just as intentional actions, one must have knowledge-how and exercise it. Thus, exercising knowledge-how is the trait we were looking for; a common characteristic of intentional actions and intentional omissions.

In the last subsection, we said that to have a capacity to act for reasons, one must

have a capacity to know how to do the action. Likewise, to have a capacity to not act for reasons, one must have a capacity to know how not to do the action. A need for knowing-how is in common between the two capacities that Setiya might suggest, and as we saw earlier, it provides a better explanation of our groundless knowledge of intention. Moreover, in both intentional action and intentional omission, one exercises the capacity to know-how, and in both, by appealing to this capacity, one can know what one is doing or omitting. In this regard, we can say that the capacity to know-how is the very capacity that is the source of groundless knowledge of intention.

In the following subsection, we address another issue facing RTI concerning prospective intentions. We will demonstrate the shortcomings of RTI in cases where one intends to do something in the future. Furthermore, we will illuminate that, based on PTA, there is no gap between intention and intentional action, and PRI is also applicable to prospective intention.

5.5.4. Prospective Intentions

The fourth issue facing RTI concerns one's access to prospective intentions. Setiya's account of knowledge of intention hinges on a rational capacity exercised when one is doing something intentionally and knowing what one is doing. When one is doing something intentionally, no (time) gap can be detected between one's intention-in-action and action. But, what if the person intends to do something in the future and does not do anything at this very moment? One may have a prospective intention to do something without actually doing it outright. To have a comprehensive account of knowledge of intention, Setiya needs to explain how one could also have groundless knowledge of such intentions without starting to do it. With regard to this issue, Setiya says:

“One virtue of this account is that it extends so readily to prospective intention. If the will is a capacity for non-perceptual, non-inferential knowledge of action, planning agency involves its application to the future: what one comes to know, or believe, is that one is going to ϕ . With the concept of intention, the capacity to form such beliefs by forming prospective intentions can be used to form the belief that one intends to ϕ , along with the belief about action itself.” (Setiya 2011: 195).

Here, it seems that by *the will* Setiya means the capacity to act for reasons. We can

say that based on his claim, the capacity to act for reasons forms and justifies both one's belief that one is ϕ -ing and one's belief that one is going to ϕ in the future, and it is the source of both kinds of intention. Accordingly, we can say that Setiya's argument for knowledge of prospective intentions runs as follows:

1. One has a reason to ϕ tomorrow.
2. One forms an intention to ϕ tomorrow.
3. An intention to ϕ tomorrow involves this belief that 'I am going to ϕ ' or at least 'I am going to ψ ,' when one does ψ in order to ϕ .
4. This belief is supported by knowledge-how-to- ϕ or at least knowledge-how-to- ψ when one does ψ in order to ϕ .
5. One knows that one is going to ϕ
6. One forms the belief that one has the intention to ϕ tomorrow.

The question raised here is that if 't1' is when the agent has a prospective intention to ϕ , and 't2' is the moment that she starts to ϕ , can we say that at t1 and t2, she has different intentions and beliefs regarding her actions? If so, the agent needs to exercise either the same capacity at t1 and t2, namely the capacity to ϕ for reasons, or two different capacities, namely the capacity to ϕ' (to ϕ tomorrow) for reasons at t1 and the capacity to ϕ for reasons at t2. The former cannot be the case since exercising a particular capacity cannot bring about two different results. If the same capacity at t1 and t2 be exercised, then the agent should not have different intentions and beliefs regarding her actions at t1 and t2.

If the latter is the case, given that the reasons are the same in both capacities, the difference between the two capacities (and accordingly, the difference between intentions and beliefs) lies in the difference between the two acts— in the present case ϕ' -ing and ϕ -ing. However, these two actions cannot be different. If we consider them as two different actions, the prospective intention to ϕ' at t1 will never be fulfilled, since what the agent will do at t2 is ϕ , not ϕ' . When one intends to do something, one is committed to fulfilling the content of one's intention. If what the agent does at t2 is not the very action she intended at t1, then she has not fulfilled her commitment.

Given the difference between prospective intention and prediction, an unwavering commitment exists throughout the interval t1 to t2. An agent who intends to do something in the future does not predicate that this action will be performed. Instead, the agent has an active role in carrying out this action and finds herself committed to fulfilling her intention. Thus, it is false to say that at t1, the agent has a different

intention from the one at t_2 and exercises various capacities. Consider t_3 as a moment between t_1 and t_2 . Suppose the agent intends at t_1 to ϕ at t_2 . During this interval, the agent is watchful to intentionally do or omit something at t_3 if, for fulfilling her commitment, she finds it necessary. So, we can say that throughout the interval until the agent fulfils her commitment, she is doing ϕ or doing ψ in order to ϕ or at least she is supervising operations to ϕ . In other words, from the beginning that the agent intends to ϕ until the very last moment that she stops doing it, she is exercising her capacity to know how to ϕ , and we can say that she is ϕ -ing.

Consequently, we can say that exercising the capacity to know-how bridges the gaps between the three different concepts that the term 'intention' can employ, and Anscombe counts them as prospective intention, intention-with-which, and intention in action⁴⁴, and as she says, "To a certain extent the three divisions of the subject ... are simply equivalent. That is to say, where the answers 'I am going to fetch my camera', 'I am fetching my camera' and 'in order to fetch my camera' are interchangeable as answers to the question 'Why?' asked when I go upstairs." Therefore, even if we sometimes use the future tense for describing our intentions, it does not mean that we cannot describe our intention in the present tense and say, 'I'm doing ...'.

In the following subsection, we will address the last difficulty facing RTI concerning the puzzling transition that occurs in Transparency of intention. The question is how one knows one's own intention by knowing one's own intentional action. We will explain that this transition has roots in the unique character of practical knowledge and that PTI can dispel the difficulty.

5.5.5. Transparency Transition

The fifth issue facing RTI concerns the transition from IV. to VI.⁴⁵. The answer cannot

⁴⁴ Anscombe has an interesting paragraph about this issue:

"A man can be doing something which he nevertheless does not do, if it is some process or enterprise which it takes time to complete and of which therefore, if it is cut short at any time, we may say that he was doing it, but did not do it. This point, however, is in no way peculiar to intentional action; for we can say that something was falling over but did not fall (since some thing stopped it). Therefore we do not appeal to the presence of intention to justify the description 'He is Y-ing'; though in some cases his own statement that he is Y-ing may, at a certain stage of the proceedings, be needed for anybody else to be able to say he is Y-ing, since not enough has gone on for that to be evident; as when we see a man doing things with an array of wires and plugs and so on." § 23, p39.

⁴⁵ The answer cannot be that compliance with the epistemic rule makes the transition possible. The question is why we stipulate such an epistemic rule.

be that compliance with the epistemic rule makes the transition possible. The question is why we can stipulate such an epistemic rule.

Although Setiya tries to explain this transition by appealing to Evans' idea, it is still unclear how the procedure explains one's access to one's own intention through knowledge of one's own intentional action. Appealing to Evans' idea, in the represented format, and saying that we can prefix the content of our propositional knowledge of intentional action with the operator "I intend . . .", appears to suggest a link between knowledge of intention and language. However, presumably, Setiya was not drawn on Evans' remark because of its possible linguistic insight, and it is hard even to think that any linguistic insight was intended by Evans himself when he was talking about knowledge of how things perceptually appear. So, if language does not play a critical role here, what else can explain this transition? In other words, what gives the agent this privilege to prefix "I intend . . ." to a judgment about what he is doing?

A possible answer to this question could be an inference—when one knows what one is doing, one can infer what one's intention is and say, "I intend . . ." However, we saw that knowledge of intention is groundless and cannot be based on an inference. Hence, something else should make this link.

This study believes that the answer lies in the unique character of practical knowledge. We mentioned that contrary to theoretical knowledge, practical knowledge, in the sense that Anscombe borrows from Aquinas, is "the cause of what it understands." In this regard, in contrast to the object of theoretical knowledge derived from the fact known, the object of practical knowledge cannot be traced back to what is or is not the case in the world. Instead, the object of practical knowledge is to be fulfilled in the world.

Practical knowledge is the cause of what it understands in the sense that not only one knows one's own intentional action, one carries out also the very action that is the object of one's knowledge. Accordingly, when one knows what one is doing intentionally, the very knowledge discloses what one is to do. Hence, such knowledge has two aspects: the first aspect of this knowledge detects 'what one is doing,' and the other aspect determines 'what one is to do.'

Determining what is to be done is a matter of practical reasoning. In this regard, explaining the notion of practical reasoning might be of some help here. Anscombe rightly says that understanding practical knowledge first requires understanding practical reasoning. (Anscombe 1957: 57). In practical reasoning, one is seeking an

answer to the question 'what do I have to do?'. According to Anscombe, practical reasoning involves something wanted, which is "at a distance from the immediate action", and by reasoning practically, the agent calculates what to do to achieve the thing wanted. (Anscombe 1957: 79). However, this calculation does not remain at the mental level without making any "movement towards" the thing wanted. As Anscombe says, an account of practical reasoning is of interest because "it describes an order which is there whenever actions are done with intentions"; otherwise, if it "were supposed to describe actual mental processes, it would in general be quite absurd". (Anscombe 1957: 80). To highlight the non-mental dimension of practical reasoning, Anscombe emphasises that "the primitive sign of wanting is trying to get" (Anscombe 1957: 68). In this respect, contrary to the conclusion of theoretical reasoning, which is the mental state of belief, the conclusion of practical reasoning is an immediate action.

However, for practical reasoning leads to an action, and there be a movement towards the thing wanted, the agent must be committed to carrying out the action. If theoretical reasoning brings a commitment to staying in line with what is or is not the case in the world, practical reasoning brings a commitment to carrying out what one has to do. In other words, the first sign of wanting is that the agent intends to attain the thing wanted. As we mentioned in the previous subsection, there is no gap between intention and action, and by intending to do the action wanted, the agent commences her movement towards the thing wanted. Thus, practical reasoning results in action and drives an intention to do the action.

Given that practical reasoning drives an intention by which the agent begins her movement towards the thing wanted, we can say that the same reasoning initiates the exercise of pertinent knowledge-how. The agent's movement towards the thing wanted cannot be arbitrary. In her reasoning, she should also consider the suitable means of attaining the thing wanted and deploy the means to achieve her end.

So far, we can conclude that practical knowledge involves a kind of practical reasoning⁴⁶ that drives one's intention to carry out the object of practical knowledge.

⁴⁶ Anscombe says: "what Aristotle meant by practical reasoning certainly included reasoning that led to action, not to omissions" § 33-34

However, according to Clarke: "There are reasons for omitting or refraining as well as for acting, and they are reasons of the same general kind: practical reasons, considerations bearing on what to do, on whether to act in one way or another. One can omit or refrain for such reasons, just as one can act for them. Omitting or refraining can be commanded ("Don't move!") or otherwise required, and such a command or requirement can be obeyed or fulfilled or, alternatively, disobeyed or flouted." (Clarke 2012b: 128)

For this reason, practical knowledge is the cause of what it understands. Given the fact that the object of practical knowledge is an action, which at the same time is the conclusion of practical reasoning, we can better understand Evans' remark and why we can prefix the content of our propositional knowledge of intentional action with the operator "I intend . . .". According to McDowell:

"The content of an intention in action is given by what one would say in expressing it, or what one would say in stating the practical knowledge one has in executing it, which comes to the same thing. And the appropriate form is 'I am doing such-and-such'." (McDowell 2010: 417)

Due to the unique character of practical knowledge, the content of one's knowledge of one's own intentional action is systematically correlated with one's intentions, and "we should insist that there is knowledge in intention only if what is happening is what one says is happening when one says what one is doing." (McDowell 2010: 430). This explains why one's intention is transparent to what one is doing intentionally.

6. Conclusions & Outlooks

The privilege of first-person knowledge is ... really more like the knowledge of a person driving a car as opposed to that of her passenger. The passenger may very well see where the driver is going, but still does not know in the immediate *executive* sense of the driver herself.

(McGeer 1996: 505)

Abstract

One specific conclusion we can make based on our current research is that Anscombe's "why?" question can be replaced by the "what?" question, which seeks the knowledge-how exercised by the agent. The most obvious finding to emerge from this study is that knowledge-how plays a crucial role in doing something intentionally, and accordingly in knowing one's own intentional action and intention. One of the strengths of this dissertation is that it examines its core theory, PTA, in different contexts and situations. Although the research has successfully demonstrated how one knows one's own intention in a privileged way, it has certain limitations in terms of explaining self-knowledge of other mental states like belief, desire and emotion. Further work needs to be done to establish whether the Phronetic rule of transparency can be extended to other mental states.

6.1. Introduction

This thesis aimed to explain one's privileged access to one's own intention. To this end, we chose the method of transparency. In the first chapter, we explained the problem of self-knowledge and reviewed the literature on transparency in the works of Richard Moran, Mathew Boyle, and Alex Byrne.

In the second chapter, we discussed intentional actions. The reason for discussing intentional actions was the connection that exists between intention and intentional action. There, we talked about the importance of knowledge-how in doing something intentionally and introduced PTA.

The similarity between self-knowledge and knowledge of intentional actions, and in particular the conceivable link between knowledge of intention and knowledge of intentional action, led us to discuss knowledge of intentional action in chapter three. There, we argued that knowledge in intentional action is a distinct type of knowledge, and we tried to explain how such distinct knowledge is possible. In other words, we said that, according to Anscombe, one's knowledge of one's own intentional action is non-observational, and we explained how one can know that one is going to do something without observing what one is bringing about in the external world. In that chapter, we turned to knowledge-how to address the problem of knowledge of intentional action.

In chapter four, we evaluated Moran's, Boyle's, and Byrne's versions of the transparency account of intention and demonstrated their difficulties in providing an epistemology of intention. In addition, we pointed out how PTA can highlight their shortcomings and sometimes help them overcome these challenges.

In chapter five, we introduced Setiya's version of transparency of intention. Based on Setiya's account, we built our own account of transparency of intention. According to our own account, if one has the capacity to know how to ϕ , and based on this capacity, knows that one is ϕ -ing, at the same time and by the same capacity, one can ascribe the intention to ϕ to oneself.

The transparency account of self-knowledge of intention that we have articulated through this dissertation could certainly be developed further in a number of significant respects. Therefore, in this final chapter, we want to consider a few of the many prospects there are for further research on the philosophy of action and the

problem of self-knowledge. In the following sections, we first present a general conclusion of our thesis. Then we will point to a specific conclusion we can make based on what we said in the fifth chapter. We suggest that we can replace Anscombe's "why?" question with the "what?" question that looks for the knowledge-how exercised by the agent. Then we come up with a recommendation for future work. We briefly present an agentic account of self-knowledge in which it seems possible to think that transparency of intention can explain self-knowledge in general.

6.2. Conclusion

The primary objective of this dissertation was to elucidate how we have privileged access to our own intentions. This research successfully developed a novel Transparency account of self-knowledge of intention, which explains the unique access individuals have to their own intentions. According to our proposed Transparency account, knowledge of intention is transparent to knowledge of intentional action. We demonstrated that the exercise of knowledge-how serves as the connecting element between these two forms of knowledge and the intentional action itself.

A key finding from this study is the critical role knowledge-how plays in performing an action intentionally and, consequently, in understanding one's intentional action and intention. One of the strengths of this dissertation lies in the examination of its core theory, the PTA, across various contexts and situations. From the ITA and its confrontation with critics to the explanations of knowledge of intention offered by Moran, Boyle, Byrne, and Setiya, we have shown that the PTA can contribute to the existing literature and provide valuable solutions to complex problems and emerging challenges.

While the research has successfully demonstrated how one has privileged access to one's own intentions, it does possess certain limitations when it comes to explaining self-knowledge of other mental states such as beliefs, desires, and emotions. Further exploration is required to determine whether the Phronetic rule of transparency can be extended to encompass other mental states. In Section 6.4, we briefly introduced a perspective suggesting that the transparency of intention might explain one's privileged access to other mental states. Prior to that, in Section 6.3, we discussed an imprecise interpretation of Anscombe's remarks on the "why?" question, and proposed the possibility of substituting Anscombe's "why?" question with a "what?" question which is looking for the knowledge-how exercised by the agent.

6.3. Why "Why?" not "What?"

In this section, we discuss the application of the "why?" question. It seems that Setiya's appeal to "the capacity to act for reasons" as a rational capacity, that is, the source of groundless knowledge of intention, originates from an interpretation of the question "why?" in which the role of this question is reduced to a device for finding reasons. Based on this reading, if one has reasons to do something and has a capacity to act in accordance with these reasons, what one is doing is intentional. In what follows, we investigate whether this question is simply a question about reasons to act or whether it has another function.

We mentioned that Setiya's recourse to the capacity to act for reasons has its roots in Anscombe's account of intentional action. In Anscombe's account, doing an action for reasons and, accordingly, the applicability of the question 'why?' is a criterion for identifying intentional actions. However, the 'why?' question, like the "how?" question, in comparison with other question words such as 'who?', 'which?', 'what?', and 'where?', asks for more explanatory and interpretative answers and is a more intricate question. To answer the "why?" question, in its special sense suggested by Anscombe, one needs to engage in a kind of psychological activity to provide reasons for what one is doing. However, we are not always able to provide an exact reason or even a reason at all to the "why?" question. Anscombe recognises cases in which, being questioned about motivations of action, one may respond something like 'I had no particular reason', without that action being unintentional. However, Anscombe does not consider the 'for no particular reason' answer as a hindrance to her account. According to her, these answers are not understood as systematically relevant to the specific question she identifies: "... it must not be supposed that because they can occur that answer would everywhere be intelligible, or that it could be the only answer ever given." (Anscombe 1957: 34).

Furthermore, in the other cases where the 'why?' question receives a positive answer, it is not certain that the answer is the actual reason for which the agent did the action. The agent would give an arbitrary answer that is not really the correct answer to the question. For instance, when someone is at her desk and writing down her ideas before falling asleep, in replying to 'why are you writing down your idea?' she can give multiple answers, which all show she has a reason for doing so. The person can say, 'I'm writing my thesis', 'my professor expects me to give him

something', 'I will forget what is in my mind if I don't write them down', 'a revolution is underway in Iran', or 'I'm hungry'. All of them could be her answer to the 'why?' question, and all of them are consistent with Anscombe's positive answers— they come under forward-looking-motives and motives-in-general. Some of them seem relevant, some not. But even the last two answers, which seem more irrelevant, might be the real motivations for the agent's action. It could be possible that whenever the agent is hungry at night, she could write better, or when she is writing something, she forgets her hunger.

Furthermore, in some cases, especially when the agent's answer is classified under Anscombe's motive-in-general, after the 'why?' question, the agent tries to find an answer just to show the rationality of her action as much as possible. The agent does not reject that there must be a reason, but she might be unconscious of that reason at the moment of doing the action. The 'why?' question brings some reasons to our consciousness. Imagine that the agent is drinking tea while she is writing. She intentionally drinks tea but does not have any conscious reason for it; she is just drinking. Precisely after the 'why?' question, the agent tries to give a reason: "I'm thirsty!" or "drinking tea can keep me awake." Even in cases where the agent's answer is under the forward-looking-motive, like when writing something to write her thesis, usually before the question, she had not considered the reason. She was just writing her idea down and had not thought about the reason, even if there must be a reason.

Consequently, there are questions raised concerning the competent person who can assess someone's answers, or the condition under which one's answers would be counted as correct. In cases where the agent cannot provide a particular reason, or perhaps in most of the other cases, it seems that only a psychologist could be the qualified person who should answer the question 'why?'. Nevertheless, it seems rational to think that Anscombe was aware of these issues, and she had a different purpose in mind regarding the "why?" question. She was not looking for an answer that has with itself psychologist's endorsement. We can see that for her, giving an exact answer to the "why?" question, namely the exact reason for the action, is not important at all. Rather, what we can find based on the book *Intention*, but has been neglected by most of the readers, is that what is essential concerning the "why?" question for Anscombe is practical reasoning — the question "why?" does not hunt for one's reason, but it traces whether there was a particular kind of reasoning behind what the agent did.

We mentioned in chapter three that an action could fall under different

descriptions, and a man may do a thing intentionally only under one description. In § 23 of her book *Intention*, Anscombe asks if there are different descriptions of what one is doing, “which is the description of an intentional action?” (Anscombe 1957: 37). Regarding her well-known example, pumping water, one can ask, “what is this man doing?”, is he pumping water, making noise, instituting the Kingdom of Heaven on earth, or else? The man, in this story, is doing all of them, and the question “what is this man doing?” (hereafter the “what?” question) is enquiring about the description under which he is doing the action intentionally. According to Anscombe, the question ‘Why?’ Can help with the “what?” question in the way that “our enquiries into the question ‘Why?’ enable us to narrow down our consideration of descriptions of what he is doing to a range covering all and only his intentional actions. ‘He is X-ing’ is a description of an intentional action if (a) it is true and (b) there is such a thing as an answer in the range I have defined to the question ‘Why are you X-ing?’” (Anscombe 1957: 37). Accordingly, in the case that someone is pumping water, if she is making noise and interrupting another person, the “why?” question can help us to discern whether she is interrupting the other person intentionally or not. If it was not intentional, then this description is put aside, and we will have a narrower list to pick out the description of intentional action.

However, in response to the question ‘Why are you X-ing?’, our list not always shrinks, but sometimes it can even expand. In reply to the question “why?”, one may say ‘to Y’, and we can ask this question repeatedly regarding Y-ing and what comes after that. X-ing, Y-ing, and all the other actions which fall into this series are the descriptions satisfying criteria (a) and (b), and hence are descriptions of intentional actions. According to Anscombe, practical reasoning and her order of ‘Why?’ questions can “be looked at as a device which reveals the order that there is in this chaos.” (Anscombe 1957: 80). But how?

In the previous chapter, we mentioned that practical reasoning involves something wanted, which is “at a distance from the immediate action”, and by reasoning practically, the agent calculates what to do to achieve the thing wanted. The first premise of a practical syllogism expresses what is wanted, and the conclusion is what one has to do to attain the thing wanted. Objects of wanting can be revealed by asking ‘What do you want?’ or the question ‘why?.’ According to Anscombe, the question ‘What do you want?’ “is the question, ‘With a view to what are you doing X, Y and Z?’ which are what he is doing.” (Anscombe 1957: 63). In this regard, both ‘What do you want?’ and the question ‘why?’ reveal what is wanted when one is doing X, Y,

and Z. However, the object of wanting was not the thing Anscombe was looking for. She was searching for the description of intentional action. Therefore, the object of wanting is only important as long as it can help us with the determination of X, Y, and Z. Therefore, 'What do you want?' and the question 'why?' are the devices for controlling whether X, Y, and Z are the conclusions of practical reasoning.

When one is X-ing, one's reasoning has resulted in doing X. X can be a direct means to achieve the thing wanted or a means to Y. If X is a means to Y, we can consider either Y is the very thing wanted because Y is the goal itself or Y is the direct or indirect means to achieve the goal. The latter means that the agent may want to Y to attain something else that is also wanted. This series of actions must stop somewhere, and that is where the action ends directly in achieving the final goal. Until then, all the actions we do in order to attain what is desirable can be desirable themselves. Therefore, we can say that sometimes practical reasoning involves other practical reasoning. (To attain the desirable thing, we have to Z; to Z (that is desirable now), we have to Y; to Y (that is desirable now), we have to X.)

In the previous chapter, we indicated that practical reasoning is present in practical knowledge, in the sense that the object of non-observational knowledge of intentional action is the very conclusion of practical reasoning. In this regard, given that intentional action is the conclusion of practical reasoning, and the agent has already answered 'what do I have to do?' when she is doing something intentionally, it seems appropriate to consider that only those descriptions that the agent has non-observational knowledge of them and be presented in her reply to the "what?" question are the descriptions of intentional action. This is because, in response to this question, the agent reveals what she intends to do and the knowledge-how she is exercising. In other words, the question "what?", in this special sense, asks only for the description of intentional action, and since the agent has already answered this question during practical reasoning, she can promptly answer it again and present an accurate description of her intentional action. In this regard, it is possible to reformulate Anscombe's assumption regarding the "why?" question as follows: An intentional action is an action to which a certain sense of the question 'what?' is given application; the sense is of course the one in which the answer, if positive, shows non-observational knowledge of the action.

6.4. Can we extend the Phronetic Rule of Transparency of Intention to other mental states: An idea for future research

In the first chapter, where we discussed Boyle's account, we briefly touched upon the Uniformity Assumption. Boyle presents the Uniformity Assumption as the idea that "a satisfactory account of our self-knowledge should be fundamentally uniform, explaining all cases of 'first-person authority' in the same basic way." (Boyle 2009: 141). As Byrne (2018) exemplifies, Ryleanism and simple inner-sense theory are unified theories of self-knowledge. According to Ryleanism, "for any mental state M, the account of how I know I am in M is broadly the same: by observing my behavior." (Byrne 2018: 16) According to a simple inner-sense theory, "I know that I am in M by focusing my "inner eye."" (Ibid.). Even philosophers like Moran, who do not provide a unified account of self-knowledge, as Boyle attempts to demonstrate and as we described in the first chapter, provide a basic explanation of one's access to one's own mental states that encompasses different mental states such as belief, desire, and emotion.

Now, the question we wish to raise is: can the Phronetic Rule of Transparency of Intention be extended to other mental states? While we do not intend to answer this question conclusively, we would like to briefly highlight two possibilities regarding the application of PRT to other mental attitudes:

1. One possibility is to follow Setiya's approach, suggesting that an individual with a specific rational capacity could have self-knowledge by adhering to a transparency rule. Both the relevant rational capacity and transparency rule could vary for different types of mental states. In this case, we would only have an explanation for knowledge of intention, and for other types of mental attitudes, we would need to identify the pertinent rational capacity and transparency rule separately.
2. Another intriguing possibility is to propose that not only our intentions, but all our mental states are transparent to our actions. For example, if someone is thirsty and has a desire for drinking water, she would typically drink water. If someone is a Christian believer, she would normally attend a church, while a

Muslim believer would go to a mosque. As rational beings, our actions (assuming sincerity) should be consistent with our mental states. For instance, if someone is sad, she would typically not engage in dancing.

Clearly, there are several challenges associated with this perspective. Not all mental states are manifested in action, and there may be conflicting mental states, such as desires that a believer considers to be sinful. However, one reason to entertain the possibility that our mental states could be transparent to our actions is the agentalist view proposed by some philosophers, such as McGeer. McGeer presents his agentalist perspective as follows:

The view I propose involves putting special emphasis on our own agency by recognizing that we are actors as well as observers and so can be good, even excellent, "predictors" of our future behavior because we have the power to make these "predictions" come true. Put simply, we are able to ensure a fit between the psychological profile we create of ourselves in first-person utterances and the acts our self-attributed intentional states are meant to predict and explain simply by adjusting our actions in appropriate ways. (McGeer 1996: 507)

According to McGeer, we typically act in ways that are consistent with our mental states, and self-knowledge depends on our role as agents. In his perspective, one's judgment about one's mental states commits one to act in ways that make sense given that mental state. He believes that this characteristic is exclusively first-personal:

I cannot make it the case that you behave in ways coherent with what I say you hope, desire, or fear any more than I can make it the case that the world is a certain way by announcing how (I think) it is; but I can and do govern my own actions in ways that fit with the claims I make about myself. (Ibid.)

Another motivation for considering the possibility that our mental states could be transparent to our actions is the account of self-knowledge that expressivists propose. In expressivist accounts, one's avowal, such as "I'm in pain," directly expresses one's self-knowledge. Bar-On says:

On our proposed account, a full explanation of the privilege must recognize avowals as expressive performances, which can be taken to reveal directly the subject's present mental condition. (Bar-On and LONG 2001)

If we could argue that our actions are, in some sense, avowals of our mental states, then we could claim that a comprehensive explanation of privilege must acknowledge actions as expressive performances. These performances can be seen as directly revealing the subject's current mental state.

As stated earlier, our intention here is not to delve deeply into the possibility of

applying PRT to other mental attitudes. We merely wanted to suggest that, in light of the agentalist and expressivist accounts of self-knowledge, exploring such a possibility could be a valuable direction for future research.

Bibliography

- Anscombe, G. E. M. 1957. *Intention*. Cambridge: Harvard University Press.
- . 1982. Medalist's Address: Action, Intention and 'Double Effect'. *Proceedings of the American Catholic Philosophical Association* 56: 12-25.
- . 2005. The Causation of Action. In *Human life, action and ethics: essays by GEM Anscombe*, ed. M. Geach and L. Gormally, 89-108. Exeter: Imprint Academic.
- Aucouturier, Valérie. 2021. Human action. In *The Anscombean Mind*, 333-54. London: Routledge.
- Audi, Robert. 1986. Intending, intentional action, and desire. In *The Ways of Desire*, ed. J. Marks, 17-38: Precedent.
- . 2011. *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. Third edition ed. New York: Routledge.
- Bar-On, Dorit. 2004. *Speaking My Mind: Expression and Self-Knowledge*. OXFORD: Clarendon Press.
- Bar-On, Dorit and DOUGLAS C. LONG. 2001. Avowals and First-Person Privilege. *Philosophy and Phenomenological Research* 62: 311-35.
- Bennett, Jonathan. 1981. Morality and Consequences. *The Tanner Lectures on Human Values* 2: 45–116.
- Boghossian, Paul A. 1989. Content and self-knowledge. *Philosophical Topics* 17: 5-26.
- . 2008. Epistemic Rules. *The Journal of Philosophy* 105: 472-500.
- Borgoni, Cristina. 2018. Basic self-knowledge and transparency. *Synthese* 195: 679-96.
- Boyle, Matthew. 2009. Two Kinds of Self-Knowledge. *Philosophy and Phenomenological Research* 78: 133-64.
- . 2011. II—Transparent Self-Knowledge. *Aristotelian Society Supplementary Volume* 85: 223-41.
- . 2019. Transparency and reflection. *Canadian Journal of Philosophy* 49: 1012-39.
- Bratman, Michael. 1984. Two Faces of Intention. *The Philosophical Review* 93: 375-405.

- Bratman, Michael E. 1987. *Intention, plans, and practical reason*. Cambridge Mass. ; London: Harvard University Press.
- Brueckner, Anthony. 2003. Two transcendental arguments concerning self-knowledge. In *New Essays on Semantic Externalism and Self-Knowledge*, ed. S. Nuccetelli, 185-200. Cambridge Mass. ; London: MIT Press.
- Burge, Tyler. 1996. Our Entitlement to Self-Knowledge: I. Tyler Burge. *Proceedings of the Aristotelian Society* 96: 91-116.
- Byrne, Alex. 2005. Introspection. *Philosophical Topics* 33: 79-104.
- . 2011. Transparency, belief, intention. *Aristotelian Society Supplementary Volume* 85: 201-21.
- . 2018. *Transparency and Self-Knowledge*. Oxford: Oxford University Press.
- Cassam, Quassim. 2011. I—Knowing What You Believe. *Proceedings of the Aristotelian Society* 111: 1-23.
- . 2014. *Self-Knowledge for Humans*: Oxford University Press.
- . 2017. What asymmetry? Knowledge of self, knowledge of others, and the inferentialist challenge. *Synthese* 194: 723-41.
- Clarke, Randolph. 2010. Intentional omissions. *Noûs* 44: 158-77.
- . 2012a. Absence of action. *Philosophical Studies* 158: 361-76.
- . 2012b. What is an omission? *Philosophical Issues* 22: 127-43.
- Coliva, Annalisa. 2016. Varieties of self-knowledge. In *The Varieties of Self-Knowledge*, 51-76: Springer.
- Cottingham, John G. 1993. *A Descartes Dictionary*. Cambridge, Massachusetts: Blackwell Reference.
- Davidson, Donald. 1963. Actions, Reasons, and Causes. *The Journal of Philosophy* 60: 685-700.
- . 1971. Agency. In *Agent, Action, and Reason*, ed. A. Marras, R. N. Bronaugh and R. W. Binkley, 1-37: University of Toronto Press.
- . 2001. *Essays on Actions and Events*. Oxford: Clarendon Press.
- Descartes, Rene. 1982. *Principles of philosophy: Translated, with explanatory notes*. Trans. V. R. M. a. R. P. MILLER. 1982 ed. Dordrecht: Kluwer Academic.
- Descartes, René. 1984. *The philosophical writings of Descartes*. Trans. R. S. John Cottingham, and Dugald Murdoch. Vol. II. New York: Cambridge University Press.
- . 1985. *The philosophical writings of Descartes*. Trans. R. S. John Cottingham, and Dugald Murdoch. Vol. I. New York: Cambridge University Press.

- . 1991. *The philosophical writings of Descartes*. Trans. R. S. John Cottingham, and Dugald Murdoch. Vol. III. New York: Cambridge University Press.
- . 1996. *Meditations on First Philosophy: With Selections from the Objections and Replies*. Trans. J. Cottingham. Cambridge: Cambridge University Press.
- Donnellan, Keith S. 1963. Knowing what I am doing. *Journal of Philosophy* 60: 401-09.
- Edgley, Roy. 1969. *Reason in Theory and Practice*. London: Hutchinson.
- Evans, Gareth. 1982. *The varieties of reference*. Edited by J. McDowell. Oxford: Oxford University Press.
- Falvey, Kevin. 2000. Knowledge in intention. *Philosophical Studies* 99: 21-44.
- Fernández, Jordi. 2013. *Transparent Minds: A Study of Self-Knowledge*: Oxford University Press.
- Gallois, André. 1996. *The world without, the mind within : an essay on first-person authority, Cambridge studies in philosophy*. Cambridge ; New York: Cambridge University Press.
- Garcia, J. L. A. 1990. The intentional and the intended. *Erkenntnis* 33: 191 - 209.
- Gertler, Brie. 2011a. *Self-Knowledge*. London: Routledge.
- . 2011b. Self-Knowledge and the Transparency of Belief. In *Self-Knowledge*, ed. A. Hatzimoysis, 125–45. Oxford: Oxford University Press.
- Ginet, Carl. 2004. Intentionally Doing and Intentionally Not Doing. *Philosophical Topics* 32: 95-110.
- Hampshire, Stuart. 1982. *Thought and Action* New Edition ed. Notre Dame, Indiana: University of Notre Dame Press.
- Harman, Gilbert. 1976. Practical Reasoning. *The Review of Metaphysics* 29: 431-63.
- Harman, Gilbert 1965. The Inference to the Best Explanation. *The Philosophical Review* 74: 88-95.
- Langton, Rae. 2004. Intention as Faith: Rae Langton. *Royal Institute of Philosophy Supplement* 55: 243-58.
- McDowell, John. 2010. What is the Content of an Intention in Action?? *Ratio* 23: 415-32.
- McGeer, Victoria. 1996. Is "Self-Knowledge" an Empirical Problem? Renegotiating the Space of Philosophical Explanation. *Journal of Philosophy* 93: 483-515.
- McLaughlin, Brian P. and Michael Tye. 1998. Is Content-Externalism Compatible with Privileged Access? *The Philosophical Review* 107: 349-80.
- Mele, Alfred R. 2012. Intentiona, Unintentional, or Neither? Middle Ground in Theory and Practice. *American Philosophical Quarterly* 49: 369-79.

- Mele, Alfred and Steven Sverdlik. 1996. Intention, intentional action, and moral responsibility. *Philosophical Studies* 82: 265 - 87.
- Moore, G. E. 1993. *Selected writings, International library of philosophy*. London: Routledge.
- Moran, Richard. 1997. Self-Knowledge: Discovery, Resolution, and Undoing. *European Journal of Philosophy* 5: 141-61.
- . 2001. *Authority and estrangement : an essay on self-knowledge*. Princeton, N.J.: Princeton University Press.
- . 2003. Responses to O'Brien and Shoemaker. *European Journal of Philosophy* 11: 402-19.
- . 2004. Précis of authority and estrangement: An essay on self-knowledge. *Philosophy and Phenomenological Research* 69: 423–26.
- Moran, Richard and Martin J. Stone. 2011. Anscombe on expression of intention : an exegesis. In *Essays on Anscombe's Intention*, ed. A. Ford, J. Hornsby and F. Stoutland. Cambridge: Harvard University Press.
- Newman, Lex. 2019. Descartes' epistemology.
<https://plato.stanford.edu/entries/descartes-epistemology/#SelfKnow>.
- O'Brien, Lucy. 2003. Moran on Agency and Self-Knowledge. *European Journal of Philosophy* 11: 375-90.
- Owens, David. 2003. Knowing Your Own Mind. *Dialogue* 42: 791-98.
- Paul, Sarah K. 2014. The transparency of mind. *Philosophy Compass* 9: 295-303.
- Paul, Sarah K. 2009. How We Know What We're Doing. *Philosophers' Imprint* 9: 1-24.
- . 2015. The transparency of intention. *Philosophical Studies* 172: 1529-48.
- Schwenkler, John. 2011. Perception and Practical Knowledge. *Philosophical Explorations* 14: 137-52.
- Searle, John R. 1980. The intentionality of intention and action. *Cognitive Science* 4: 47-70.
- Setiya, Kieran. 2007. *Reasons without rationalism*. Princeton, N.J.: Princeton University Press.
- . 2008. Practical knowledge. *Ethics* 118: 388-409.
- . 2009. Practical Knowledge Revisited. *Ethics* 120: 128-37.
- . 2011. Knowledge of intention. In *Essays on Anscombe's Intention*, ed. A. Ford, J. Hornsby and F. Stoutland, 170-97. Cambridge: Harvard University Press.
- . 2012. Transparency and Inference. *Proceedings of the Aristotelian Society* 112: 263-68.

- Shoemaker, Sydney. 1988. On Knowing One's Own Mind. *Philosophical Perspectives* 2: 183-209.
- . 2003. Moran on Self-Knowledge. *European Journal of Philosophy* 11: 391-401.
- Smith, Kurt. 2015. *The Descartes Dictionary*. London: Bloomsbury Publishing.
- Thompson, Michael. 2011. Anscombe's Intention and practical knowledge. In *Essays on Anscombe's Intention*, ed. A. Ford, J. Hornsby and F. Stoutland, 198–210. London: Harvard University Press.
- Velleman, J. David. 1991. Practical Reflection. *Ethics* 102: 117-28.
- . 2000. The Possibility of Practical Reason. *Philosophical Studies* 121: 263-75.
- Vermazen, Bruce. 1985. Negative acts. In *Essays on Davidson: Actions and Events*, ed. B. Vermazen and M. B. Hintikka, 93--104: Oxford University Press.
- Walton, Douglas N. 1980. Omitting, Refraining and Letting Happen. *American Philosophical Quarterly* 17: 319 - 26.
- Way, Jonathan. 2007. Self-knowledge and the limits of transparency. *Analysis* 67: 223-30.
- Winokur, Benjamin. 2021. Critical Reasoning and the Inferential Transparency Method. *Res Philosophica* 98: 23-42.